Training and Simulation

Proceedings of the HFES Europe Chapter

Annual Meeting in Dortmund, November 1994

Karel Anton Brookhuis
Clemens Weikert
Carl Richard Cavonius

University of Groningen   Traffic Research Centre

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
16 November 1994

3. REPORT TYPE AND DATES COVERED
Conference Proceedings

4. TITLE AND SUBTITLE
Training and Simulation—Proceedings of the HFES Europe Chapter Annual Meeting in Dortmund, November 1994

5. FUNDING NUMBERS
F6170894W0756

6. AUTHOR(S)
Conference Committee

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Universitat Dortmund
Ardeystr. 67
Dortmund D-44139
Germany

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
EOARD
PSC 802 BOX 14
FPO 09499-0200

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
CSP 94-1045

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Final Proceedings for Human Factors in Training and Simulation, 7 November 1994 - 8 November 1994

The Topics covered include: cognitive performance, simulator based training, and 3D radar displays for air traffic control problems.

14. SUBJECT TERMS

15. NUMBER OF PAGES
139

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19, SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

Training and Simulation

Proceedings of the HFES Europe Chapter annual meeting in Dortmund
November 1994

Edited by:

Karel Anton Brookhuis

Clemens Weikert

Carl Richard Cavonius

CIP-gegevens Koninklijke Bibliotheek, Den Haag

Training and Simulation - Karel Brookhuis, Clemens Weikert & Richard Cavonius (Eds.) - Traffic Research Centre, University of Groningen - ill.

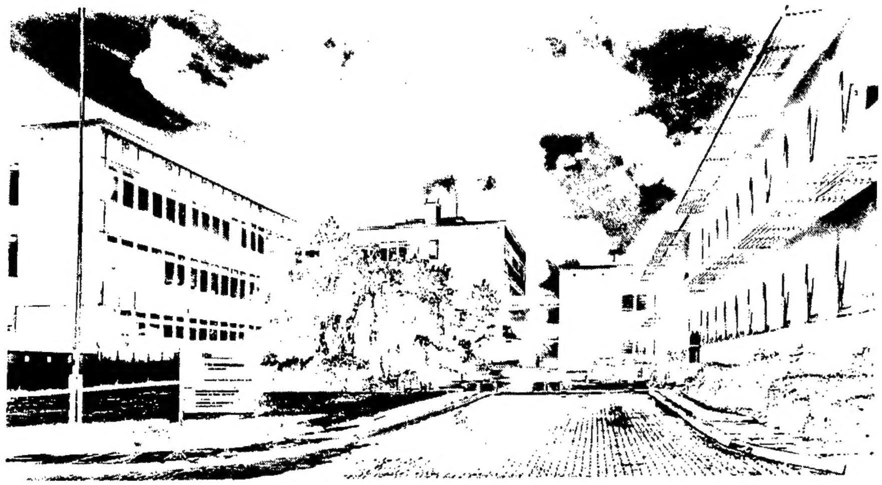
Proceedings of the Human Factors and Ergonomics Society / Europe Chapter annual meeting in Dortmund, November 1994.

With references.

ISBN 90-6807-319-2

Keywords: ergonomics, training, simulation, safety

Contents



The *Institut für Arbeitsphysiologie an der Universität Dortmund* hosted the 1994 meeting of the Europe Chapter of the Human Factors and Ergonomics Society.

Preface

This is the fourth in the series of Proceedings of the (HFES) Europe Chapter's Annual Scientific Meetings, based on the papers presented in Dortmund on the 7th and 8th of November 1994. In 1991 the Executive Council of the Europe Chapter decided to encourage publication of a Proceedings Volume from each Scientific Meeting, in order to enhance interest and attendance.

The theme of the meeting in Dortmund was Training and Simulation, a subject of rapidly growing interest. The opening paper is a written version of the dinner speech by Dr. Johan Riemersma, expending on the theme of the meeting, followed by ten papers that were presented and accepted as manuscripts in this Proceedings, some of them adapted slightly by us to conform to the style of the booklet. Not uncommon to Proceedings, the endproduct contains contributions of varying quality, to be judged by the reader, but we decided to leave most of what was written intact. We are grateful to the contributing authors of this Proceedings and want to thank them particularly for their patience and willingness to revise the manuscripts in line with our comments.

We owe special words of gratitude to the "Institut für Arbeitsphysiologie an der Universität Dortmund", that hosted the Meeting. Also, we are grateful to Dr. Dick de Waard for his dilligent editorial work. Finally, we wish to thank the United States Air Force European Office of Aerospace Research and Development for its contribution to this Proceedings and the success of the meeting.

The editors

Current Issues in Training and Simulation

Dinner speech to the Europe Chapter of the Human Factors and Ergonomics Society

Johan Riemersma
TNO-HFRI Soesterberg
The Netherlands

The Industrial Research and Development Advisory Committee of the European Commission (IRDAC) has recently (Anonymous, 1994) issued a Report on Quality and Relevance, with sub-titles: *the challenge to European education*, and: *unlocking Europe's human potential.*

A main theme in the report is the cooperation between industry and education that is needed to meet the challenge of providing a strong capacity for innovation and quality in order for Europe to survive in the international competition.

As main threats are identified: firstly, the underestimation of both the need to change and the speed of adaptation required, and secondly, the low awareness of the educational system of its central responsibility for equipping young people with the relevant knowledge, skills and attitudes to address this challenge of providing a strong capacity for innovation and quality.

To tackle these threats seven main areas for action were identified, four of which I will describe briefly.

1 Developing total competence in people

Total competence is the mix of knowledge, skills, personal abilities and attributes of a person. Total competence is stressed, because traditionally the educational system tends to focus and put highest value on the acquisition of formal knowledge only. Responsibility for skill acquisition was primarily delegated to the work environment, while attitudes and values were seen as resulting mainly from family life and society at large. IRDAC stresses the need for cooperation and communication skills as main attributes. IRDAC also stresses the need for more proper matching of course programme objectives with changing employment requirements. This implies firstly that companies should be more explicit about the broad competence needs of their work-force and secondly that (vocational) education policy should resist the temptation of narrow specialisations.

2 Preparing people and society for a lifetime of learning

IRDAC stresses the need for developing learning abilities instead of just acquiring certain certifications (qualifications). Formal education should be explicitly designed as preparatory for continuing education and training and thus also explicitly address learning skills. Society expenditures should be more evenly balanced over initial and continuing education.

3 Adopting quality concepts in education and training

Here, IRDAC advocates a dual approach in stating firstly the requirement that quality concepts should be part of the content of curricula and secondly by requiring the assessment of the quality of education and training themselves. A measure of quality is not provided, however.

4 Stimulating a learning culture in companies

Central is the concept of a 'learning organisation'. In my view this is a misleading term since learning is not a faculty of organisations but only of individuals, be it in cooperation. Of crucial importance for learning organizations is creating an atmosphere and setting in which people can learn from mistakes in a productive way instead of learning to conceal responsibilities and to divert blame.

Of the other areas, I want to mention a few remarks.

Interesting is the call for more emphasis in initial education on developing scientific and technology literacy. Relevant for the subject of this paper is also the statement that there is a strong need for multidisciplinary research on training and education.

I have repeated these points somewhat extensively, but the main message of IRDAC is that firstly the human factor plays a crucial role and could be, or perhaps even hás to be, the decisive competitive edge for Europe for the decennia to come, and secondly that the quality of the human factor is a function of education and training; the latter should be much more geared to the challenges of innovation, flexibility and adaptiveness.

This challenge has to be confronted with current practices of education and training. I don't want to undervalue many attempts to innovate our educational and vocational training systems but the mainstream can still be characterized as essentially cottage industry. The prototype learning situations are the classroom and the mentor-learner situation in which the burden of 'finding out instruction strategies and tactics and creating challenging learning environments' lies on isolated, individual teachers and mentors, often poorly prepared for these important tasks and usually not provided with the proper means to overcome their inherent limitations.

You may think by now that I am getting too serious for just a dinner speech. So it is the moment for a break. In this break I will tell you something about metallurgy and the secret of Damascene swords (Maugh, 1982).

It has been claimed that Alexander the Great already carried weapons of Damascus steel as long ago as 320 B.C. but this has not been proven. It is certainly true however that such weapons were in use from the beginning of the Islamic period A.D. 620 until well after the Dark Ages. They are named after Damascus because there they were first encountered by Europeans (and Europeans are quick in imposing names from their own perspective), but the metal (or alloy) really originated in India and in raw form was called *wootz*. The weapons forged from this metal were famous for their exceptional toughness and their unrivaled retention of a cutting edge and hence they dominated warfare for centuries.

Think now a moment about the analogy between an educational system processing children as raw material into our work-force and the blacksmith forging the *wootz* as raw material into swords.

The surprising thing about Damascene swords is two-edged.

The first fact is that the Europeans could not work properly with *wootz*. It just crumbled under their efforts to forge it. To justify their failure, where 'barbarians' could, they built a layer of mystique around the metal. This is not to say they didn't try or tried it only half-heartedly. The efforts to unravel the secret continued even in the 18th and 19th century and even Michael Faraday took part in the effort and in the process almost discovered stainless steel. Only in our days the crucial process variables of carbon content and forging temperatures were identified and this was a side finding of research into superplastic metals by Sherby and Wadsworth of respectively Stanford University and the Lockheed Palo Alto Research Laboratory.

The high carbon content of *wootz* prevented forging at the usual high temperatures of the European ovens; it had to be forged at a much lower temperature. (700-900 degrees as opposed to the European standard for forging of 1300 degrees; at this temperature *wootz* is partly solid and partly liquid and hence falls a part in forging).

Nobody ever thought about changing habitual forging temperatures or experimenting with this variable.

The second fact is that the way of cooling the worked blade of a Damascener sword is important. Some Persian texts insist that the red-hot blade should be quenched by plunging it into the belly of a muscular Nubian slave presumably thereby (by killing the slave) endowing the sword with a spirit of strength. Other texts suggest a more humane recipe of cooling in the urine of a red-haired boy or, alternatively, of goats that had eaten only ferns for the preceding three days.

For me this story illustrates by analogy very clearly the pitfalls of preconceptions and habits in the thinking about educational and learning processes on the one hand and the survival of ill-founded concepts and strategies, demonstrated to be successful, on the other hand. Moreover, given the complexity of most learning and training situations, such pitfalls are hardly avoidable.

From the Dark Ages again to the present. We see a great surge of new technologies in the realms of training and education. The information technology and the new avenues of communication provide a shrinking world in the sense that you can almost talk with everybody else in a moment-to-moment fashion and in principle you have access to all the knowledge in the world. Training can be done in simulated situations in which you never realise that they are not real. Long haul networks now connect simulators in USA, South Korea, Germany, the Netherlands, the UK and so on and collective training of armed forces is made possible for units geographically wide apart. Computer-based training, Computer-managed instruction, Simulators, Virtual Environments and Distance learning are the emerging ways of individualizing, delivering and timing training and instruction and these concepts can be made more and more intelligent by using techniques of AI.

There is a very strong technology push in changing the ways for developing and delivering training and instruction and still the education and training community seems not really ripe for it. As in learning you obviously cannot make too big steps. And still that is just what is needed. You don't need computers to present you with textbook education or talking heads. Computers can be used for providing rich learning environments but without any navigational aids you lose track of what you were supposed to learn from them. The dilemma always is that for any new skill or topic you want to acquire you need guidance to acquire it in a more effective way than just trial-and-error. Learning needs guidance and dialogues and is thus in principle a social affair of acquiring shared knowledge and representations you can talk about.

In the late sixties we had the "School is Dead" movement. I will not discuss the underlying philosophies, but use their analysis of the social functions of schools (schools in the American sense, encompassing all educational institutions of all levels, thus also including universities).

The four social functions identified (Reimer, 1971) were:

- custodial care, freeing parents for work outside the home
- social-role selection by processes of selection and routing
- indoctrination
- education, i.e. imparting knowledge and skills.

The last function, education, was shown to take about 20 % of the time of teachers, which were mainly occupied with behaviour control, custodial care and administrative routines. It is the combination of these functions which makes

schooling expensive, educationally inefficient and, I like to add, very resistant to change.

The title of my talk was "Current issues in Training and Simulation" and so I will discuss them in a few more words, using the building blocks I have cut out so far.

The main emerging message of IRDAC was that firstly, the human factor plays a crucial role and is the decisive competitive edge for Europe for the decennia to come and secondly that the quality of the human factor is a function of education and training, the latter should be much more geared to the challenges of innovation, flexibility and adaptiveness.

Given the multiple roles of our educational systems, it is unlikely that they will change to the degree and with the speed IRDAC seems to require. Most computers introduced in the educational institutions are used for administrative purposes even when they were intended for developing computer literacy (in students, not in teachers!) whatever that may be. The content of curricula is still decided upon in esoteric ways in committees and is in no way derived from a sound analysis of what people need to learn for firstly getting decent jobs and secondly for being able to cope with modern rapidly-changing job demands.

The emerging technologies have yet to prove their versatility. A main problem is the degrading of the richness of communication between pupil and teaching entity; all the subtle communications of voice intonations and body language signals are not yet implemented in mouse control and could well lead to loss of contact between learner and teaching device and thus to loss of motivation to learn. The failure of "programmed instruction" should be a warning that a too narrow-minded approach on the basis of intrinsically sound principles cannot succeed. Yet, the emerging technologies enable a much enriched learning environment with episodic experiences, simulations of all kinds, much more varied presentations of exemplars of concepts, a challenging way of presenting problem and so on. But learners have to <u>learn</u> how to make best use of the new learning environments and they have to be <u>motivated</u> for it.

From a Human Factors standpoint much can be said about educational systems. I come back to the IRDAC statement that there is a strong need for multidisciplinary research on training and education and I add to it the observation that there is a large gap between traditional providers of education and training on the one hand and the developers of advanced training and instruction technology as computer-based-instruction and simulations on the other hand. The former are not very aware of the limitations of their teaching strategies and often model their activities according to their own experiences in a school system, while the latter often have no choice but to 'copy' these deficient teaching strategies in a new form; otherwise their technology will not be accepted. This leads to a head-on collision and competition for the societies (shrinking) resources for education and training and not to a dialogue about the best way to make new win-win combinations.

It is my personal judgement that the Human Factors specialist could play a very important, facilitating and catalysing role as mediator in a necessarily multi-disciplinary effort of innovating the ways we train and educate people. My claim for this role is mainly based on the observation that the Human Factors specialists generally have their roots in both Social Sciences and in the field of technology applications and that they usually already work in multi-disciplinary settings.

I see three main ways in which Human Factor specialists can contribute substantially:

- to advocate the systems approach to training and education
- to contribute to system ergonomics and interfaces
- to translate and apply findings in such basic sciences as cognitive and learning psychology and pedagogy etc, to training and education concepts using new technologies.

IRDAC stresses the need for cooperation and communication skills because it is quite clear that nobody works in isolation and that individual knowledge and skills are not sufficient to attain goals at the group level. Working in teams is stressed and thus skills enabling to do just that have to be developed. This is quite in contrast with the implicit priorities of most educational systems, which are tailored for acquiring individual competencies. Research on collaborative learning and team training is now emerging and could contribute to further meeting also this requirement.

I have not addressed many of the current issues in training and simulation but instead I have tried to communicate a more global view on the emerging technologies in education and training and ways to innovate our delivering systems. Such innovation is a challenge that can only be answered by cooperative, multi-disciplinary efforts in which Human Factor specialists can play a crucial role.

References

Anonymous (1994). *Quality and Relevance*, IRDAC. Brussels.
Maugh, Th. H. (1982). A metallurgical tale of irony, *Science*, *215*, 153.
Reimer, E. (1971). *School is Dead.* Penguin Educational Specials.

Spatial visuo-motor compatibility and manual control in a tracking task

W.H. Ehrenstein, C.R. Cavonius & E. Lewke
Institut für Arbeitsphysiologie an der Universität Dortmund
Abteilung Sinnes- und Neurophysiologie, Dortmund
Germany

Abstract

We asked how spatial compatibility between target and tracking directions affects tracking performance. The subject's task was to keep a small target within the center of a 0.6-deg window defined by two parallel bars. The angular correspondence between target and tracking directions varied in steps of 45 deg from compatible (0 deg) to incompatible (180 deg) and intermediate (perpendicular and diagonal) arrangements. In Experiment I, the trajectory of the target had a constant orientation while the orientation of the tracking rail was varied. In Experiment II, the tracking rail remained constant while the orientation of the visual display was varied. In both experiments, tracking performance (time on target, root-mean-square error) was found to vary with angular visuo-motor compatibility, with a performance minimum at 180 deg in Experiment I and at 225 deg in Experiment II. This effect was strongest for untrained subjects but persisted even after practice.

Introduction

The principle of compatibility was originally introduced in the context of human factors during World War II in an effort to enhance signal detection. A visual display was added to an auditory display and, interestingly enough, research on this dual-modality display showed that it was not always advantageous to have additional visual information. Detection thresholds typically increased when the display provided "incompatible" attributes. This occurred, for instance, when the auditory stimulus varied in amplitude, while the visual signal varied in spatial position. Compatibility would have required that both signals varied in, for example, intensity in a congruent way. This result, which was first presented in 1951 by Arnold Small in England at a meeting of the Ergonomics Research Society, attracted the attention of Paul Fitts who described the compatibility principle as "a landmark of great significance with broad applicability" (Small, 1990).

During the next few years, Fitts and his co-workers applied the compatibility principle not only to stimulus-stimulus (S-S) pairings but also to stimulus-response (S-R) and to response-response (R-R) compatibility. Today, relationships between stimulus and response properties dominate this area, especially in ergonomics.

Most studies on S-R compatibility address problems such as spatial coding, human information processing and motor performance, man-machine interaction, and optimal design of displays or keybords (see Proctor & Reeve, 1990; and Wickens, 1992, for reviews). A similarly impressive number of studies exist in the field of manual tracking (see Poulton, 1974; Knight, 1987). However, possible relationships between S-R compatibility and manual tracking seem so far to be largely neglected. The present study attempts to bridge the gap between compatibility and tracking research. Specifically, we asked whether the relationship between the direction of stimulus movement and direction of tracking would affect manual control and tracking performance. Two experiments were performed. In Experiment I, the visual input was kept constant while the orientation of tracking was varied. In Experiment II, the direction of tracking remained constant while the orientation of the visual input was varied.

Method

The signal was a small spot of light that moved in a straight line with sinusoidal acceleration on a computer display. It started at the center of the screen and moved, beginning either to the left or right, through five complete cycles. The amplitude of motion was 16 deg and the average velocity 3.3 deg/s, so that each trial lasted approximately 48 s. The visual signal had to be tracked by moving a stylus along a rail on a digi-pad (GENIUS: HiSketch 1212; sampling rate 67 data pairs/s; accuracy 0.2 mm) which lay horizontal on a desk between the subject and the display screen. Both the computer screen and the response board could be rotated and the angle between stimulus and response was measured in degrees, with horizontal for the display and the subject's fronto-parallel plane for the tracking arbitrarily taken as the reference or 0 deg; see Fig. 1.

Stimulus motion and recording of tracking behavior were controlled by a PC (IBM 486) with a purpose-made timer-card that allowed timing with an accuracy of 0.2 msec, independent of the computer time base. Procedure and data acquisition were controlled by self-made software that allowed recording of the tracking behavior in steps of 15 ms.

Eight persons, four female and four male, aged between 18 and 34 years with normal or corrected-to-normal vision, served as subjects in Experiment I, and eight persons (four females and four males), aged between 20 and 44 years in Experiment II. Four of the same subjects participated in both experiments. All subjects were unpracticed, i.e. they had not participated before in this or similar tracking tasks. Subjects were given some practice trials and an explanation of the task before the formal experiment.

With their head fixed on a chin-and-forehead support the subject's eyes were 57 cm from the computer screen, looking straight ahead at the position of the stimulus onset. The task was to keep the target within the center of a 0.6 deg-window defined by two bars by moving a stylus along a 10-mm wide rail on a digi-pad.

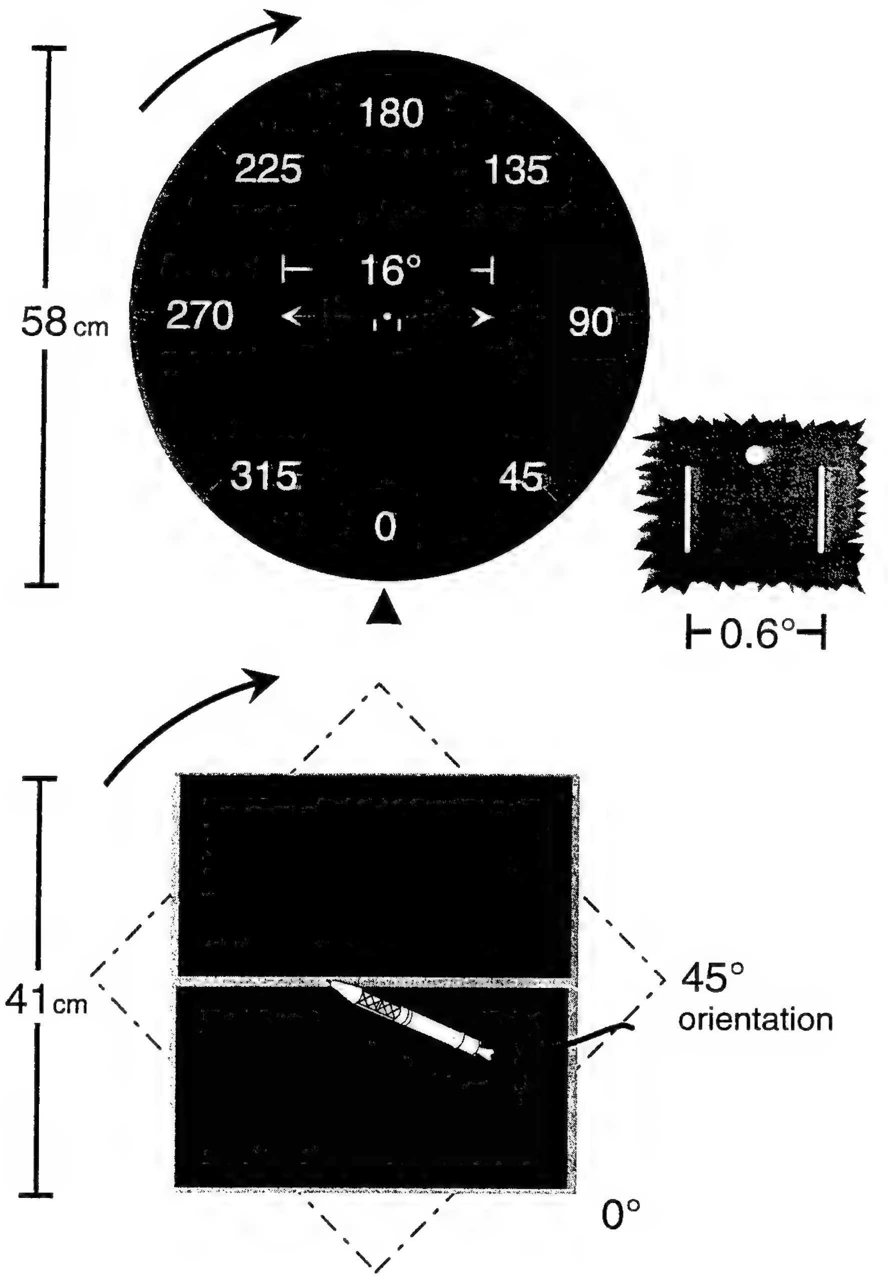


Figure 1. Schematic view of the target screen (above) and response pad (below), both shown in their 0 deg (reference) orientation. The screen was vertical, perpendicular to the subject's visual axis, while the pad was horizontal on a desk in front of the subject.

In Experiment I the target motion was always horizontal, while the tracking path was rotated in eight steps of 45 deg (clockwise as viewed from above) so that the angular S-R correspondence varied from compatible (0 deg, i.e., target motion and tracking in the same direction) through incompatible (180 deg: tracking opposite to target motion); in Experiment II the tracking was constant from left to right across the subject's fronto-parallel plane, while the orientation of the computer screen was varied in steps of 45 deg, clockwise as viewed from the subject's position (see Fig. 1).

The eight angular conditions were presented twice, with different initial directions of stimulus motion, in a balanced sequence across subjects in a single block of trials in a Latin-square design. In order to test for the effect of practice, subjects were tested in three subsequent blocks of trials with rest periods. A session lasted approximately 2½ hours.

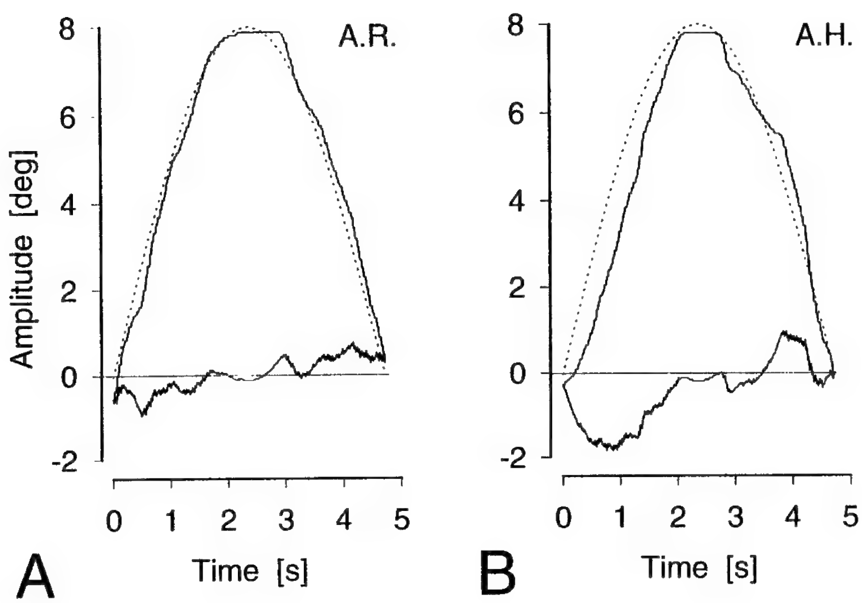


Figure 2. Space-time diagram of the recordings on one half-circle of sine-wave motion for two subjects (A.R. and A.H.). Stimulus motion (dashed line), tracking trace (continuous line) and the difference between stimulus and response (continuous line around zero-error line) are given.

Results

Figure 2 shows examples of the recordings: the space-time diagram gives the stimulus motion (dashed line) and of the tracking trace (continuous line) together with the difference between stimulus and response (lower continuous line) for one half-cycle for subject AR who gave a fairly good performance (left) and for subject AH (right) with a poorer performance.

Quantitative treatment of the tracking performance on a micro-scale, i.e., within a resolution of 15-ms steps, is reported elsewhere for some of the data obtained in Experiment I (Ganz et al., 1996; this volume).

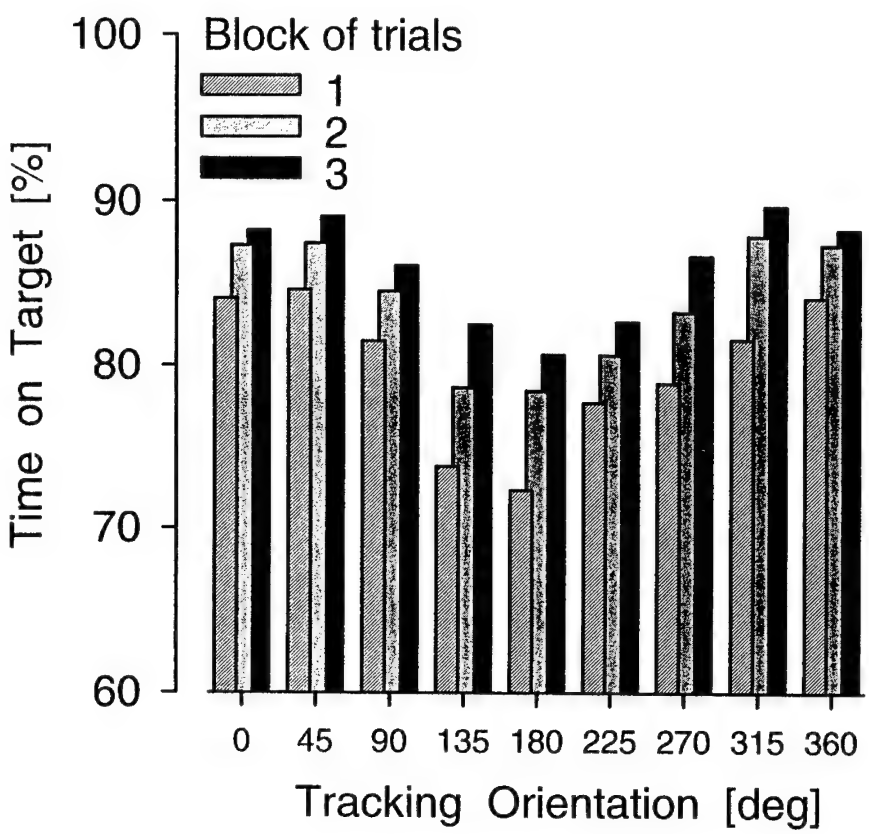


Figure 3. Mean performance (time on target) obtained in Experiment I.

Here we have taken two global measures of performance: the time on target T (the per cent of time in which the tracking was within the 0.6-deg window) and the

root-mean-square (RMS) error E (the variance of the tracking position from the midpoint of the window measured in minutes of arc).

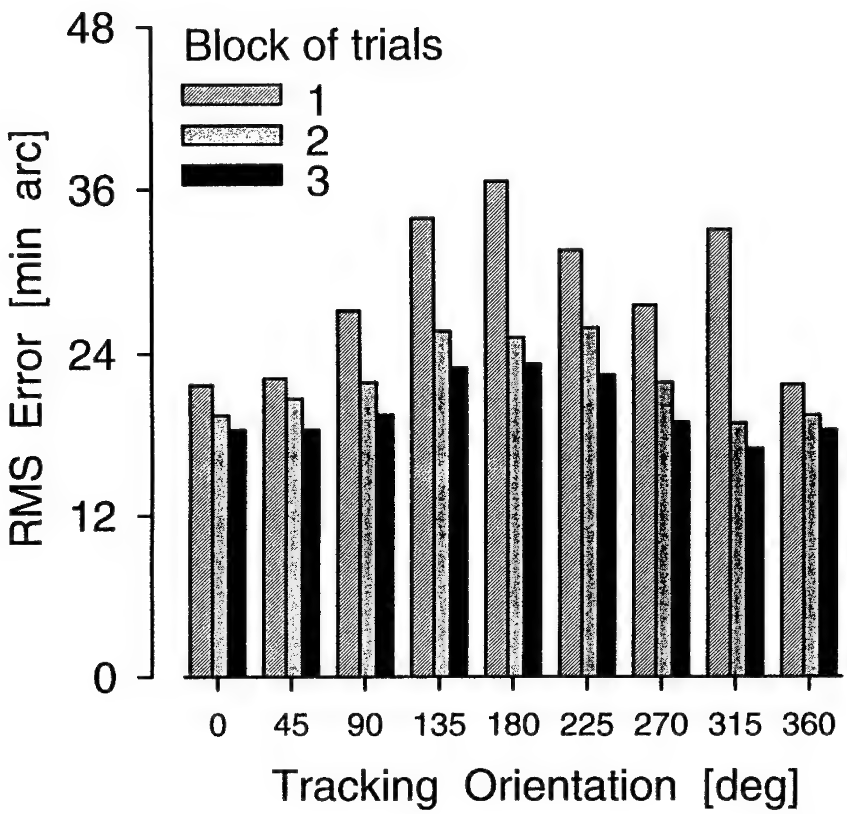


Figure 4. Mean performance (RMS error) obtained in Experiment I.

Mean performance data obtained in the three blocks of trials of Experiment I are shown in Figures 3 and 4, for time on target and RMS error, respectively. (Data of 360 deg are by definition identical with those of 0 deg). As seen in Figure 3, performance expressed by time on target systematically changes as a function of tracking orientation, with best performance at compatible orientations (0, 45, 315 deg) with a clear performance loss at incompatible orientations (135, 180, 225 deg). Performance improves with practice in the second and third block of trials in a similar manner at all orientations, i.e. the difference across orientations persists with practice. Figure 4 shows essentially the same dependency of tracking performance on tracking orientation when RMS error is taken as a measure (higher scores indicate poorer performance).

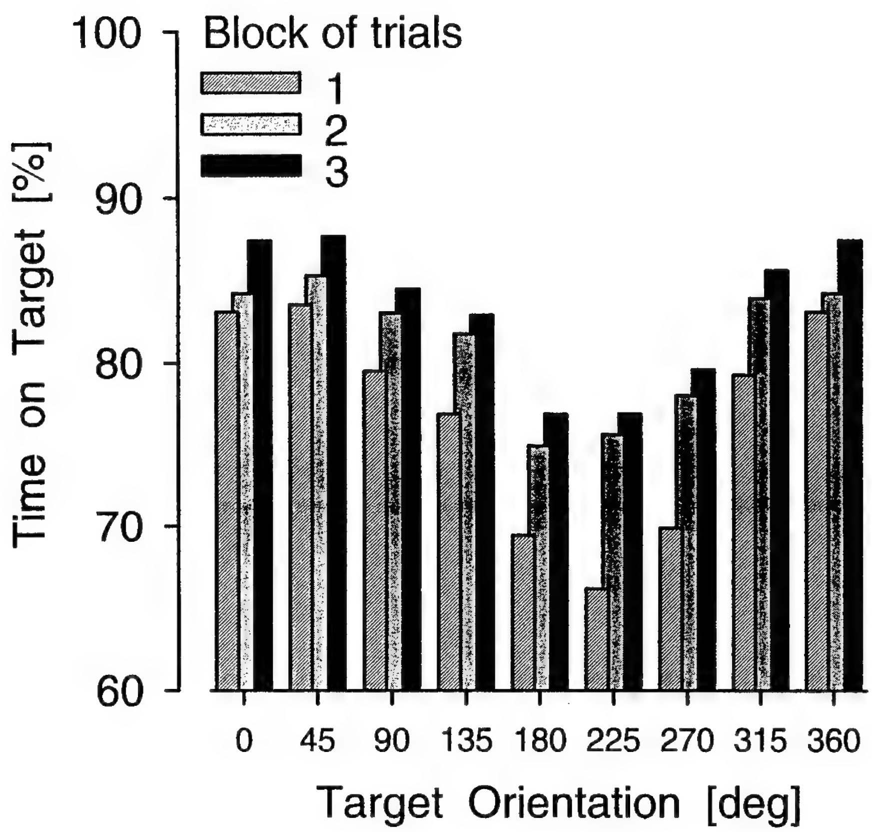


Figure 5. Mean performance (time on target) obtained in Experiment II.

Figures 5 and 6 show mean performance data obtained in the three blocks of trials of Experiment II, in which stimulus orientation was varied. Again, time on target changes with the orientation of the target trajectory; however, the poorest performance is not, as in Experiment I, at 180 deg but shifted to 225 deg, i.e., towards an oblique orientation at which the target trajectory is from upper-left to lower-right. Again, performance improves with practice, but the dependence on target orientation is essentially preserved. Similarly, the RMS error shows performance minimum at 225 deg; this asymmetric dependence on orientation may be even more pronounced with practice, i.e., in the third block of trials, performance minimum is shifted further to a perpendicular target orientation of 270 deg.

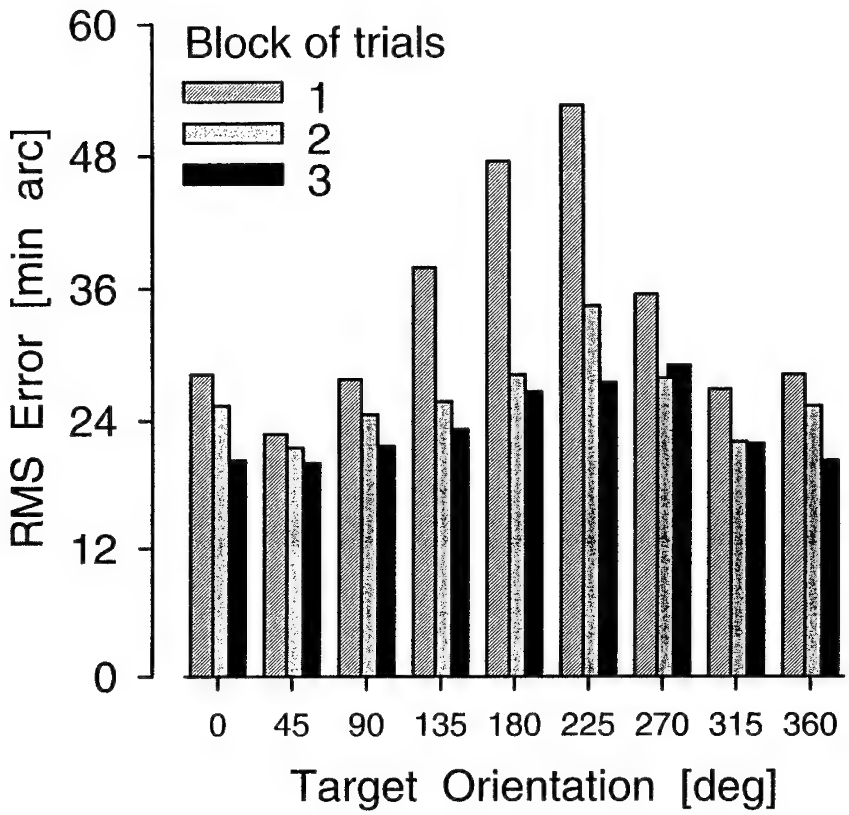


Figure 6. Mean performance (RMS error) obtained in Experiment II.

The data of Experiments I and II were subjected to a 3-way within-subjects analysis of variance (ANOVA), in which the factors were initial direction of the target motion (e.g., left, right) order of blocks of trials (3), and orientation (tracking or target) (8 steps), separately for the two performance measures (T, E); the main results are summarized in Table 1.

The initial direction of target motion was never significant. In both experiments, the order of blocks of trials, reflecting effects of practice, was significant. Both target and tracking orientation, which affect the effect of angular compatibility, were highly significant. No interactions in Experiment I were significant, whereas in Experiment II one interaction, that between initial target direction and order of blocks of trials (D x O), was significant.

Table 1: Main results of the analyses of variance.

	Performance Measure					
	Time on target			RMS error		
Factors	*F*	df	*p*	*F*	df	*p*
	Experiment I (3way ANOVA)					
Direction	2.2	1, 7	0.18	4.1	1, 7	0.08
Order of trials	26.2	2, 7	0.001	18.0	2, 7	0.004
Tracking Orientation	10.4	2.9, 20.6*	0.0002	5.9	1.7, 12.1*	0.02
	Experiment II (3-way ANOVA)					
Direction	1.1	1, 7	0.33	0.02	1, 7	0.88
Order of trials	7.4	2, 7	0.03	8.8	2, 7	0.02
Target orientation	12.5	3.3, 22.8*	0.0001	5.77	1.8, 12.9*	0.02
D x O	6.6	1, 7	0.037	0.17	1, 7	0.69
	Experiment I, II (4-way ANOVA)					
Exp. Condition	13.6	1, 3	0.035	12.3	1, 3	0.04
Orientation	5.2	2.0, 6.1*	0.04	7.1	2.1, 6.3*	0.02

* Greenhouse-Geisser adjusted df

An additional 4-way ANOVA (Tab. 1) that included the factor of experimental condition (tracking vs. target orientation) for the data of the four subjects who participated in both experiments showed that the main effect of experimental condition (orientation) was significant while otherwise essentially confirming the outcome of the previous 3-way ANOVAs.

Discussion

These results offer clear evidence that manual control in visuomotor tracking depends on angular S-R compatibility. Tracking performance decreases with

angular S-R separation, or in other words, it improves with the degree of spatial compatibility. This compatibility effect persists with practice, indicating that performance reflects task complexity rather than familiarity with the task conditions. Similar results are known to occur for angular S-R compatibility with static visual stimuli and key-press responses (e.g. Simon & Wolf, 1963; Ehrenstein et al., 1989). Thus, compatibility between display location or movement and the location or movement of the respective operator response may be a critical factor, which should be taken into account in control-system design (Wickens 1992).

Several questions remain. In Experiment I, tracking orientation was varied, which changed not only the angular S-R compatibility relationship but also the angle from the body at which the movement is made. The latter factor has been shown to affect linear pursuit movements as well (Corrigan & Brogden, 1948). To clarify this issue, further research should vary the tracking orientation together with the target orientation, so that the angular S-R compatibility is held constant while the angle of movement with respect to the body is changed. In Experiment II, the latter factor cannot account for the results, since tracking movements were always the same. It thus remains an open question why the poorest performance was not found at 180 deg orientation, i.e. for a condition where tracking and target directions were opposite to each other, but at a somewhat greater angle. A possible factor accounting for such an asymmetric dependence of performance on angular S-R compatibility might be handedness. All subjects tested here were right-handed. However, preliminary data obtained with lefthanders, and also measurements of the tracking performance when right-handers tracked with their left hands, yielded essentially the same results and hence do not support this possibility.

The present study has just opened the door to a rich field of future investigations that attempt to bridge between the two traditional branches of research on spatial S-R compatibility, and tracking skill and manual control.

Acknowledgements

We would like to thank Ludger Blanke for providing the timer card and Peter Dillmann for technical assistance in the experimental set up.

References

Corrigan, R.E. & Brogden, W.J. (1948). The effect of angle upon precision of linear pursuit movements. *American Journal of Psychology, 61*, 502-510.

Ehrenstein, W.H., Schroeder-Heister, P. & Heister, G. (1989). Spatial S-R compatibility with orthogonal stimulus-response relationship. *Perception & Psychophysics, 45*, 215-220.

Ganz, R.E., Ehrenstein, W.H., Schlykowa, L. & Cavonius, C.R. (1996). Quantifying the dynamic complexity of visuo-motor tracking performance. (this volume).

Knight, J.L. (1987). Manual control and tracking. In Salvendy, G. (Ed.) *Handbook of Human Factors* (pp. 182-218). New York: Wiley.

Poulton, E.C. (1974). *Tracking Skill and Manual Control*. New York: Academic Press.

Proctor, R.W. & Reeve, T.G. (Eds.) (1990). *Stimulus-Response Compatibility. An Integrated Perspective*. Amsterdam: North-Holland.

Simon, J.R. & Wolf, J.D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics, 6*, 99 - 105.

Small, A.M. (1990). In Proctor, R.W. & Reeve, T.G. (Eds.). *Stimulus-Response Compatibility. An Integrated Perspective* (pp. v-vi). Amsterdam: North-Holland.

Wickens, C.D. (1992). *Engineering Psychology and Human Performance (2nd ed.)* New York: HarperCollins.

Quantifying the Dynamic Complexity of Visuo-Motor Tracking Performance

R.E. Ganz[1], W.H. Ehrenstein[2], L. Schlykowa[3] and C.R. Cavonius[2]

[1]Institut für Medizinische Psychologie, Universitätsklinikum Essen, Germany
[2]Institut für Arbeitsphysiologie an der Universität Dortmund, Germany
[3]Carl-Ludwig-Institut für Physiologie, Universität Leipzig, Germany

Abstract

We introduce a method developed from the theory of non-linear dynamics that allows one to calculate the dynamic complexity (correlation dimension D) of subjects' movement patterns during a visuo-motor tracking task. In a computer-aided experiment, we measured time series (4096 data points in steps of 15 ms) of the spatial deviations of tracking movements from a sinusoidally accelerated target (0.1 Hz) and used Takens' method of time delays to reconstruct the phase trajectory from these data. Following Grassberger and Procaccia, all pairs of points on this trajectory within a small distance r from each other were summed to yield a correlation integral C(r) that, for a sufficiently high embedding dimension m ($m \geq 2D+1$), corresponds to r^D. The D-values obtained for two exemplary subjects suggest that motor training results in an increase in the dynamic complexity of the movement patterns.

Introduction

The extensive biodynamical analyses of the Russian physiologist N.A. Bernstein suggest that the major result of motor learning is the development of skills of correction rather than the refinement of stereotyped patterns of movement (Bernstein, 1988). Here the term "motor correction" tries to capture the idea that the result of the continuous interplay of action and perception is under constant revision. We introduce a new method developed from the theory of non-linear dynamics (chaos theory) that allows us to quantify the dynamic properties of these corrections.

Experiment

The subject's task was to track a light spot that moved sinusoidally along a 16-deg horizontal path at 0.1 Hz on a computer screen by moving a digi-pad stylus so as to hold the spot in the virtual centre of a 0.6-deg window. The stylus was constrained to move along a rail (width 10 mm) on a compatibly-oriented digi-pad (GENIUS: HiSketch 1212, accuracy 0.2 mm) (i.e., rightward movement of the stimulus spot required rightward movement of the stylus, and vice versa; for further details, see Ehrenstein et al., 1996, this volume). Using a custom-made timer-card (precision

0.2 ms) and software, we were able to obtain data on a micro-scale: we recorded time series (4096 data points, precision 0.04 deg) of the spatial deviations of the subject's track (centre of the window) from the target trace at intervals of 15 ms. In the following, these spatial deviations will be referred to as error data.

Data Analysis

Figure 1 shows two examples of the recordings: the upper curve represents a time series of errors of an untrained subject who had never before performed this or any similar tracking tasks, and who exhibits rather poor tracking performance (the root-mean-square [RMS] error is 0.48 deg); whereas the trace below was taken from a well-trained subject who had frequent experience with this and similar tracking tasks, and who exhibits good performance (RMS error is 0.2 deg).

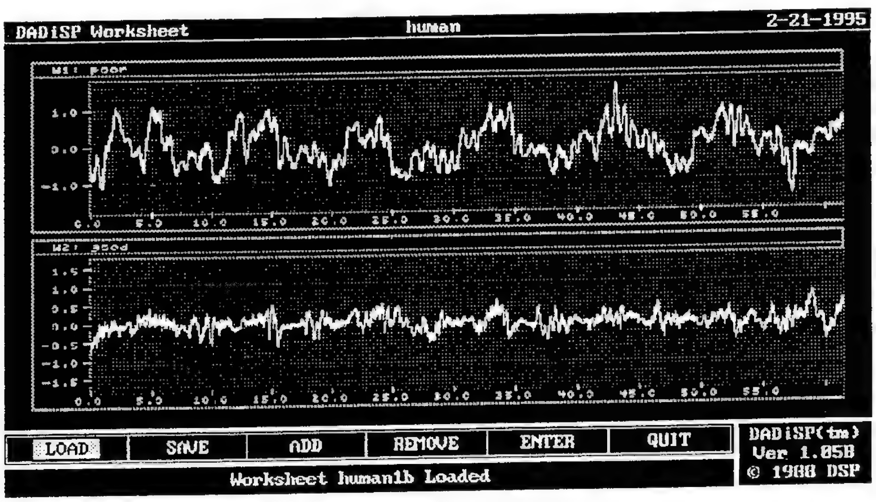


Figure 1. Time series (4096 data points in steps of 15 ms, precision 0.04 deg) of the spatial deviations of tracking movements from a sinusoidally moving target (0.1 Hz) taken from an untrained subject with a poor tracking performance (above) and a well-trained subject with a good tracking performance (below).

The dynamics of the second error signal (trained subject) appears to be more complex than that of the first (untrained subject). This first impression is qualitatively supported by the corresponding power spectra of the (z-transformed) signals (Figure 2): Although both error signals have their major component at the frequency of the target spot* (0.1 Hz), the power spectrum of the lower signal also seems to be more complex than that of the upper.

* One might argue that the fundamental frequency of the error signal originates from an internal human oscillator rather than from the external target trace. However, by systematically varying the frequency of the target, we found that the fundamental frequency of the spatial errors follows the frequency of the target.

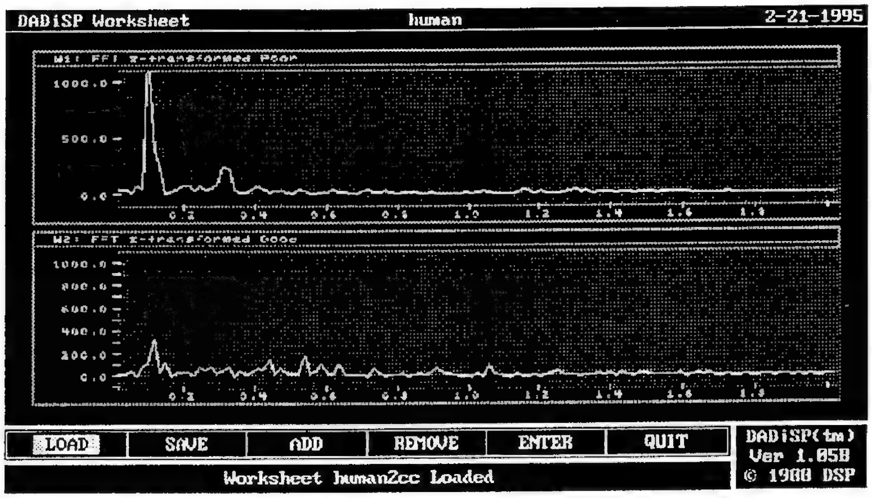


Figure 2. Power spectra of the z-transformed signals of Figure 1. The abscissa shows the relevant frequencies in Hz; the ordinate gives the power of the z-transformed signal at a particular frequency. Above: untrained subject; below: well-trained subject.

But how can we derive a quantitative measure of the dynamic complexity of these data? Here we show that one possible solution to this problem is to calculate the "correlation dimension" (Grassberger & Procaccia, 1983) of the (z-transformed) error signal. The principle idea behind the calculation of the correlation dimension is that the dynamic structure of an unknown system (in this case, the visuo-motor system) can be reconstructed from the temporal behaviour of a single output variable of this system (the spatial errors). Following Takens (1981), we constructed a "phase portrait" of vectors $X(t)$ of the (z-transformed) scalar error data x(t) in an m-dimensional phase space by taking time-delayed samples of the scalar data such that

$$X(t) = [x(t), x(t+\tau), x(t+2\tau), ..., x(t+(m-1)\tau]$$

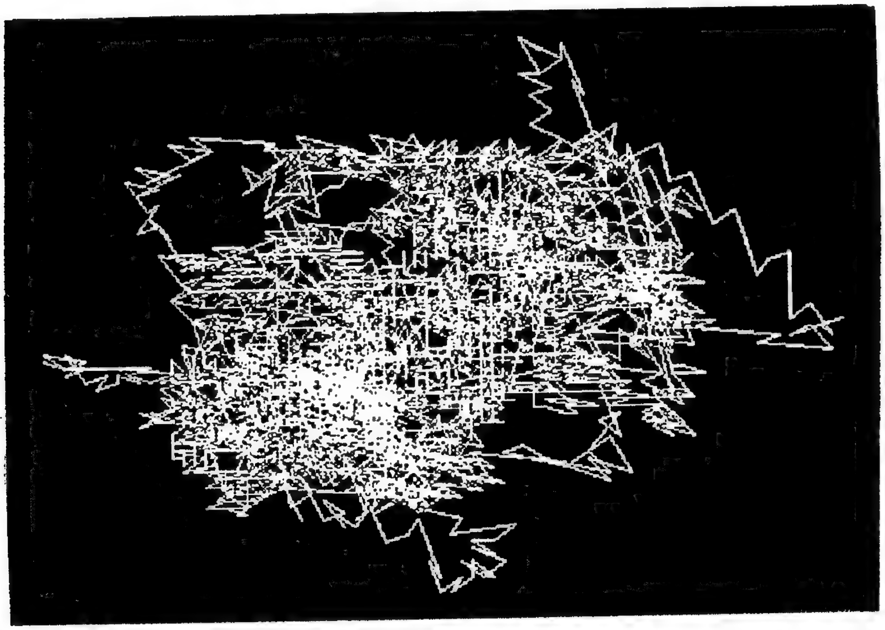


Figure 3. Two-dimensional phase portrait of the upper signal of Figure 1 (untrained subject). The abscissa shows the signal at a time t; the ordinate gives the corresponding value of the signal at a later time t+ τ (τ = 0.99 s).

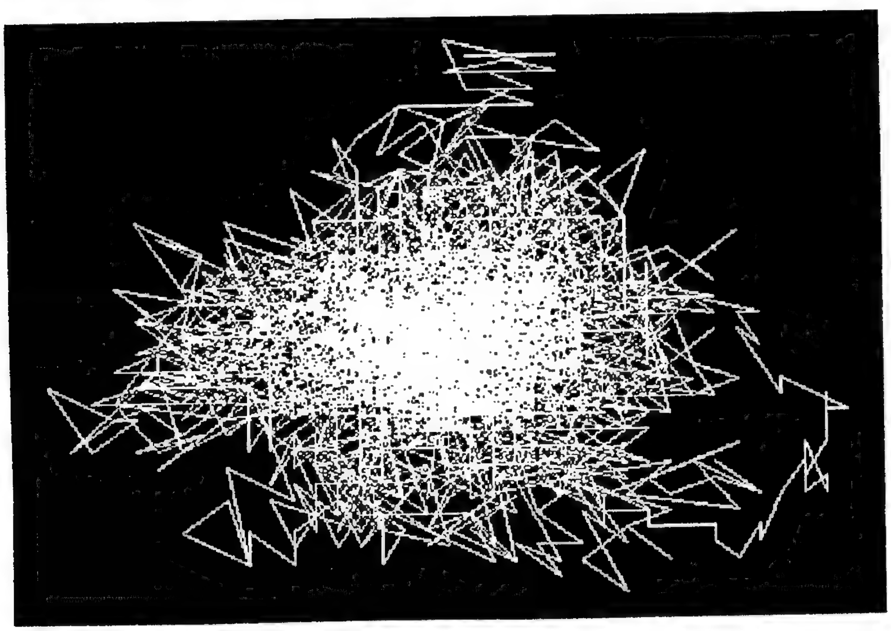


Figure 4. Two-dimensional phase portrait of the lower signal of Figure 1 (well-trained subject). The abscissa shows the signal at a time t; the ordinate gives the corresponding value of the signal at a later time t+τ (τ = 0.99 s).

Figures 3 and 4 show the phase portraits of the error data of Figure 1 in a two-dimensional reconstruction of the phase space. The time lag τ was fixed at 0.99 s (maximum decay time of the autocorrelation functions of the two signals). The topological structure of the whole cloud of points, the so-called "attractor", reflects the dynamic organization of the subject's movement. Judging by the dense "nucleus", we may expect the trained subject's attractor to have a highly complex internal structure (Figure 4).

One quantitative measure of the attractor's complexity is the "correlation dimension", D. Using the method of Grassberger and Procaccia (1983), all pairs of points on the attractor within a small spatial distance r from each other are added up in a successively higher-dimensional reconstruction of the phase space to yield a correlation integral C(r) that, for a sufficiently high embedding dimension m ($m \geq D+1$; Takens, 1981), corresponds to r^D •.

Results

Figure 5 shows the correlation dimensions D of the discussed error data for a successively higher-dimensional reconstruction of the phase space. The curves converge well. The D-values support the qualitative impressions of Figures 1-4: the movement pattern of the well-trained subject is quantitatively more complex (D = 6.75) than that of the untrained one (D = 4.86)° .

Conclusion

This study aimed at developing a quantitative measure of the dynamic complexity of movement patterns. We have introduced the correlation dimension D as such an index. The D-values obtained for two exemplary subjects suggest that motor training results in an increase of the dynamic complexity of the movement patterns.

Acknowledgements.

We would like to thank Gregor Weibels for valuable comments.

• A Gedankenexperiment may help to illustrate this idea: Start out from any point of the attractor (Figure 3 or 4), look in all directions of the embedding space and count all your neighbouring points that are at a maximum distance r. The sum of all points corresponds to r raised to the power of the dimension D.

To simplify matters, imagine that the attractor is a square with sides of length r. A square has a dimension of 2, and so the sum of all points within this square corresponds to r^2. Likewise, a cube with sides of length r has a dimension of 3, and so the sum of points wihin this cube corresponds to r^3. Now, the attractors of the error signals presented here are much more complex and irregular, so that their dimensions are fractal and greater than 3.

° One might argue that the obtained numerical difference between the D-values is not overwhelming. But note: D-values are exponents. For this reason the difference of nearly 2 between the two D-values corresponds, geometrically speaking, to the difference of complexity that exists between a point and a square or, likewise, between a line segment and a cube.

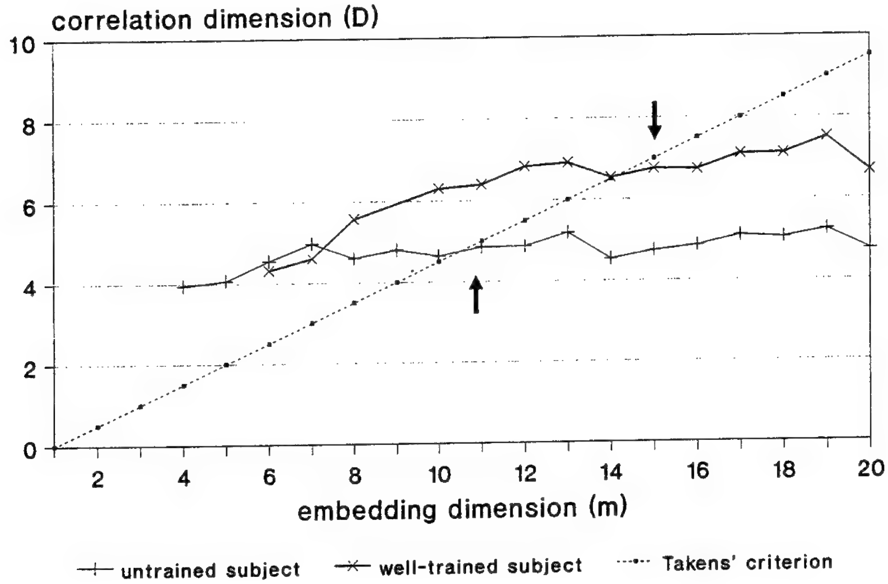


Figure 5. Correlation dimensions D of the signals of Figure 1 for successive embedding dimensions m. Arrows indicate the points with the smallest m-values that satisfy the Takens criterion ($m \geq 2D+1$).

References

Bernstein, N. A. (1988). *Bewegungsphysiologie* (2nd ed.). Leipzig, Germany: J. A. Barth.

Ehrenstein, W. H., Cavonius, C. R. & Lewke, E. (1996). *Spatial visuo-motor compatibility and manual control in a tracking task.* (this volume).

Grassberger, P. & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica, 9D*, 189 - 208.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. A. Rand and L. S. Young (Eds.), *Lecture Notes in Mathematics, No. 898: Dynamical Systems and Turbulence*. Berlin: Springer; 366 - 381.

Cognitive performance changes in a dry hyperbaric environment equivalent to 180 meters of seawater

Will A. C. Spijkers
Abteilung Arbeitspsychologie
Institut für Arbeitsphysiologie an der Universität Dortmund
Germany

Abstract

Cognitive performance changes were examined during the saturation phase (4 days) of simulated deep-sea diving (180 metres) and during the decompression phase (7 days) with four professional divers. Three tasks were investigated: a zero-order pursuit tracking with four difficulty levels (two preview and two amplitude conditions), a focussed attention reaction-time task with zero, compatible and incompatible distractor signals and a spatial working memory. The performance data showed that all tasks were impaired by the hyperbaric conditions of 19 bar. Spatial working memory was impaired as were tracking and focussed attention performance. The data suggested that high pressure impaired tracking performance largely by affecting the visual information processing component. In the focussed attention task reaction time was prolonged, but the ability to filter out response-irrelevant signals was not affected which indicates that high pressure does not lead to a general slowing of all information processing.

The results also supported the possibility of selection divers for optimal execution of certain tasks in dry hyperbaric conditions during the planning phase of a dive. However, before using the simulation tasks of present study the results should be validated with regard to real tasks and, furthermore, whether they can be generalized to different levels of pressure should be examined.

Introduction

Off-shore divers play an important role in, for example, the exploration of marine energy sources. Together with the development of saturation diving techniques and the extension of underwater operation areas the demands on divers have also increased. On the one hand the diver must master modern underwater working-techniques, on the other it is necessary to move and work in an environment for which the human body is not suited. As a consequence, the diver must meet considerable physical and psychological demands in order to guarantee a safe stay in the hyperbaric environment. In order to perform efficiently, cognitive and motor skills are required in addition to the usual diving qualifications (Zinkowski, 1978). One possibility for increasing work safety and effectiveness in the underwater workplace is selecting divers for specialized tasks during the planning phase, based on their individual cognitive and motor skills.

Several factors limit the depth to which human divers can go because these affect both their physiological well-being and performance efficiency. Pressure and gas mixture are two factors that are relatively amenable to investigation. Because of the narcotic effect of nitrogen at pressure, it is usually replaced by helium at depths of more than 50m of sea water (MSW). Although helium is considered to be an inert gas, rapid compression with oxyhelium can result in a condition known as High Pressure Nervous Syndrome (HPNS). HPNS causes dizziness, vomiting, tremors and a general performance decrement. An extensive and critical review of the behavioural effects of inert gas narcosis is provided by Fowler, Ackles and Porlier (1985). Their conclusion is that cognitive abilities, such as sentence comprehension, conceptual reasoning, immediate memory, mental arithmetic, digit cancellation, two-choice reaction tasks, card sorting and the like, are more susceptible to narcosis than manual abilities, such as the pegboard task (e.g., Shilling, Werts and Schandelmeier, 1976). However, this difference in susceptibility was questioned by Ross (1989) because the methods of measuring the performance decrements are not similar.

Lewis and Baddeley (1981) and Logie and Baddeley (1985) studied a wide range of cognitive tasks during simulated deep-sea diving (under dry pressure chamber conditions) at 61 MSW and at several greater depths, ranging from 300 to 540 MSW. They found that impairments in cognitive performance were clearly present at depths that exceed 300 MSW, whereas at 61 MSW no clear picture of performance decrement was obtained. However, other studies reported cognitive performance decrements at depths much less than 300 MSW (e.g. Biersner and Cameron, 1970; O'Reilly, 1974; 1977). Both Lewis and Baddeley (1981) and Logie and Baddeley (1985) concluded that the effects of pressure when breathing oxyhelium were not as general as predicted by the slowed processing model, but were selective because significant impairments were not observed in all aspects of cognitive functioning. Memory tasks, and cognitive tasks that put great demands on working memory and perceptual processing speed were affected more than pattern recognition and verbal reasoning tasks.

In the present study tasks were investigated that differed in several respects from those usually studied. Common to all of the tasks was that they measured speed of processing and working memory. Distinguishing features were that they involved abilities that are necessary (a) in eliminating distracting signals, as in the focussed-attention task of Eriksen and Schulz (1979), and (b) in predicting the future spatial position of a signal in tracking, thereby evaluating the spatial component of working memory. Furthermore, movement speed was examined by means of manipulating the amplitude of the tracking signal.

Simulation of hyperbaric welding

In addition to extending the type of cognitive processes investigated, the present study tried to predict performance under hyperbaric conditions. The following two aspects were considered for selecting an appropriate task. First, the selection of a representative task i.e., a typical underwater task, and second, identification of

suitable predictors of performance at surface level and under hyperbaric conditions. A central demand to be met by simulations of real systems is a functional correspondence between the two systems (Jorna & Moraal, 1988) which includes the following aspects:

a) *physical correspondence*, defining the extent of identical hardware characteristics of simulation and the real system, such as the dynamics of the motion system,

b) *behavioural correspondence*, defining the extent of similar man-in-the-loop behaviour, i.e., common cognitive and motor demands.

One of the frequently-performed underwater tasks is welding, and this task was used as target task. The goal in welding (for example a crack or fissure) consists of a line, the course of which can be either straight and regular or curved and irregular. The required action is tracking the trajectory of this line with a hand-held tool, and the task is to minimize the tracking error at a certain constant velocity. By the use of a zero-order pursuit-tracking task a high degree of functional correspondence is achieved between the chosen task of dry hyperbaric welding and the simulation task, as there is a real physical and behavioural correspondence between those two tasks. In both systems, the real and simulated one, task difficulty increases with the amplitude of the signal to be tracked, given a constant frequency of oscillation. This is analogous to the speed-accuracy trade-off principle with discrete aiming movements postulated by Fitts (1954): given constant speed of movement, the accuracy of movement decreases with increasing amplitude of movement. Another factor that influences task difficulty of both welding and tracking is the amount of preview, i.e., the degree to which the track ahead is visible and predictable. By manipulating preview the time to choose and program a corrective movement in advance is varied: the shorter the preview the later the corrective movement can be realized. As a consequence, signal-processing-speed must increase in order to still minimize error and this in turn increases task difficulty (Reid & Drewell, 1972).

In a pursuit-tracking task the following basic factors, so-called human operator limits, are commonly distinguished (Wickens, 1984): processing time, and processing resources of spatial working memory. Processing time defines the individual's time necessary for signal-processing which depends, among other things, on the individual's ability of a) processing a given signal-position, b) error-calculation and c) the programming of the required corrective movement. This factor is expressed by the effective time delay in tracking (McRuer & Jex 1967) and is analogous to reaction time with discrete aiming movements (Wickens, 1984). In order to assess the speed of processing, the reaction time task of the focussed-attention method developed by Eriksen and Eriksen (1974) was used, because this laboratory task combines both response choice and distracting signals. The compatibility of the distracting signal allows one to the examine processing ability for separating real signals from other interfering ones. Processing resources of spatial working memory concerns the processing-capacity of spatial working

memory in tracking (Gill et al., 1982) to which the individual ability of building an internal model of the tracking system is related. This in turn serves as a basis for error-calculation within the tracking process. Spatial working memory was assessed by means of a spatial memory task developed by Adam, Ketelaars, Kingma & Hoek (1993).

This type of research is extremely expensive and does not lend itself to multiple-experiment studies or to large sample sizes, but one possible solution is to test a small number of subjects at several points throughout the different diving phases, to provide a profile of performance. Questionaires that assess sleep quality and psychological states, such as emotional state and alertness, were not used, because previous studies have shown that the obtained impairments in cognitive performance were not related to changes in psychological state variables (e.g., Lewis and Baddeley, 1981; Logie and Baddeley, 1985).

When a decrement of performance is related only to a compression effect, i.e., to physiological adaptation to increased pressure, it can be expected that performance will more or less recover during the saturation phase. When the performance remains low during the saturation phase, it is unlikely that the adaptation proces alone is responsible for the observed performance decrement.

As to the effectiveness of the various predictor tasks, it was hypothesized that the tracking task itself represents the best predictor task when performance on all tasks decrease, and the correlation between tracking on the one hand and the focussed attention and spatial memory task on the other hand do not change during the saturation phase. When the correlation between tracking and the other two tasks diverge during the saturation phase, this can be taken as evidence that the component represented in the particular task that has a higher correlation has a higher predictive value.

Method

The experiment was part of the dive project ENTEX 32 of the GISMER (Groupe d'Intervention Sous La Mer) at the Centre Hyperbare of the French Navy in Toulon. The project included both wet and dry hyperbaric environments of 19 bar, which corresponds to 180 meters of seawater. Only the performance on cognitive tasks in the dry hyperbaric chamber will be reported here.

Subjects

Four experienced divers of the French Navy (30 - 38 years of age) with similar diving experience took part in the program, which involved seven pre-dive days, one compression, four isopression and seven decompression days. All subjects had medical examinations prior to the dive, and were in good physical condition. All had normal or corrected-to-normal vision.

Apparatus

All tasks were run on an IBM-compatible computer located in the control room. Tasks were presented on a black and white monitor. The image of the computer monitor was relayed to a second monitor by means of a video camera placed 15 cm in front of the screen which allowed the experimenter close supervision of the experiment. During the tasks the subjects sat in the dry hyperbaric chamber in front of the second monitor at a distance of 70 cm. This monitor was located in the control room behind a 5 cm thick glass window. A response board with two keys and two joy-sticks were positioned before the subjects. Comfortable and error-less functioning of these manipulanda under high pressure was demonstrated before the start of the experiment.

Procedure and design

Psychological cognitive performance was examined with three tests. These were a pursuit tracking, a focussed attention, and a spatial memory task. As described above the pursuit tracking task was considered to be a simulation of welding activities. The procedure of these tasks and their experimental variations are described for each task separately. Both the tracking and the focussed-attention task were controlled by the software program ERTS (Experimental Run Time System).

Pursuit-tracking task

The subject's task was to align continuously the cursor with the tracking signal by left- and rightward movements of a joy-stick. The tracking signal (filtered noise; cut-off frequencies .36 and .70 Hz) consisted of a curved upwards-moving vertical path presented with pre-defined maximal amplitude, thus creating an irregular line-signal that resembled one that occurs with natural fissures. It was presented at the center of the monitor within two horizontal and parallel lines. The horizontal lines defined a window and enabled the manipulation of the amount of preview by choosing the appropriate distance between them. On the upper horizontal line a downward pointing arrow served as cursor and the subject's task was to align the tip of the arrow with the tracking line by lateral displacement of the cursor along the line. The speed of the vertical signal movement was constant. System gain (relation between amplitude of system output and amplitude of operator input) was 1, corresponding to real-life welding. There was no system-induced time delay. The joy-stick had no mechanical supression or spring lead. Movement of the joy-stick was executed with one hand only.

Preview at the tracking trajectory was either 150 or 4,500 ms and these preview conditions were examined under two amplitude conditions: 1.75 and 4.5 cm. This gave an easy task (amplitude 1.75 cm and 4,500 ms preview) and a difficult task (4.5 cm amplitude and 150 ms preview). The remaining combinations of amplitude and preview were conditions of intermediate difficulty level. The four tracking conditions were balanced across sessions and subjects.

During a session each subject performed four tracking trials, one trial of 165 s duration for each condition. The first 15 s of a trial were a warm-up period and were discarded from the analysis. Cursor and target position were sampled at 30 Hz and the root mean square (RMS) of the deviation between target and cursor position was used as the dependent variable for the tracking performance. A trial was started by the experimenter as soon as the subject had positioned the cursor at the (initially stationary) tracking signal.

Focussed-Attention Task

This task was similar to the one used by Eriksen and Eriksen (1979). The task as used in order to operationalize the basic factor "information processing time". It consisted of a 2-choice reaction time task in which a left (right) key-press was required as soon as possible after the presentation of a capital A (B) at the center of the display. The imperative signals always appeared at the same location, 500 ms after presentation of a visual warning signal (also 500 ms) which consisted of a point which served as fixation point. The imperative signal disappeared with the initiation of the response. The Index and Middle fingers of the dominant hand were assigned to the left and right key with the fingers resting at the keys during the session. A new trial was initiated 500 ms after the response.

Four stimulus configurations were presented: single, neutral, compatible and incompatible. In the single condition either one of the letters appeared without distractor signals, whereas in the remaining three (distractor) conditions the target letter was flanked by a letter on each side. The target letter (A or B) was flanked either by letters that were identical to the target letter (compatible), or by the opposite letter (incompatible), or by neutral letters (X) not associated with an experimentally defined response (neutral condition).

Again, the four conditions represented different levels of task difficulty, according to Eriksen & Eriksen (1979) with the single condition as the easiest one, the compatible condition was less easy, followed by the neutral condition, whereas the incompatible condition was considered as the most difficult one.

A letter (height: 13 mm, width: 7.5 mm) covered about 1.0 x 0.6° of visual angle and the letters were spaced equidistantly. In the distractor conditions the total width of the stimulus configuration was 42 mm, which corresponds to 3.5° of visual angle at a viewing distance of 70 cm. The four conditions were presented in blocks of 20 trials and randomly distributed within a session with the restriction that each of the four conditions was presented an equal number of times within a session.

Spatial working memory task

In this task subjects had to bring a cursor (6 x 2.5 mm) from its home position at the center of the display to the remembered location of a target which was presented 2,000 ms before. Each trial started with the presentation of a fixation point ("+" sign; 500 ms) at the center of the screen. The target stimulus was a

single " * " sign. It appeared on the screen immediately after the disappearance of the fixation point in an area indicated by a 15.0 cm (width) and 13.4 cm (height) box constisting of strings of " * " signs (27 x 21 signs).

Within the box an imaginary 25 (horizontal) x 19 (vertical) grid created 475 cells. The target stimulus could appear in all but the middle cell (where the fixation point was positioned), thereby providing 474 possible stimulus positions. Five imaginary "rectangular zones" within the box, centered on the fixation point established five general distances. Presentation duration of the target was 100 ms and during a block of 25 trials, the targets were equally distributed over the 5 distance zones.

After disappearance of the target the screen remained dark during 2,000 ms. After this delay the cursor appeared at the fixation point and subjects had to move the cursor to the perceived position of the target by means of a joy-stick. Once it had arrived at the intended target location, subjects confirmed their response by pressing the "fire button" on the joy-stick. This responce elicited a 750 ms flickering of the correct target location and hence provided feedback regarding the accuracy of response. Movement of cursor was produced by discrete and equal steps from one position in the imaginary grid to an adjacent other one by discrete steps of joy-stick movement in horizontal or vertical direction only.

Localization performance was quantified by calculating the distance between the target and the subject's reponse location. Since the distance between adjacent horizontal and two adjacent vertical stimulus positions was 5.75 and 6.75 mm, an error of less than 5 mm represents an averaged remembered location immediately adjacent to the target position (c.f. Adam et al., 1993).

Administration of the tasks

Each diver executed all tasks under all experimental conditions. They always performed the tasks in the same fixed order: first spatial working memory, then focussed-attention and finally the tracking tasks. However, within the latter two tasks the order of experimental conditions was randomized (see the individual procedure sections). Before the experimental hyperbaric phase subjects practiced the tasks on six consecutive days, four times a day in the hyperbaric chamber. After the 24 training sessions, baseline data of all tasks were recorded in two sessions on the 7th day with the same environmental conditions as in the experimental hyperbaric phase but at normal pressure (1 bar). No tests were performed on the compression day, when the pressure was increased to 19 bar which is equivalent to 180 metres depth. Pressure was gradually increased at a rate of 7 m/hr. Two sessions were run on the following four days of saturation: one morning (10 - 12 h) and one afternoon (15 - 17 h) session. Each diver performed the tasks at about the same time of day during the four isocompression (saturation phase) and seven decompression days. Testing order for each subject was balanced between divers and sessions during the entire experimental phase. Total duration of all tests amounted to about 25 min.

Results

In Table 1 the mean results are presented for each condition of the three tasks under baseline (1 bar) and hyperbaric (19 bar) conditions.

Baseline performance

For the tracking task a two-way ANOVA with factors amplitude and preview revealed significant effects for both main factors at sea level. Increasing the Amplitude increased RSME from 1.657 to 2.897, $F(1,7) = 134.73$; $p < .001$, while shortening the preview from 4,500 to 150 ms caused a small, but significant reduction in tracking performance from 2.079 to 2.475, $F(1,7) = 8.66$; $p < .05$. The interaction was not significant ($p = .71$). In the Focussed attention task a t-test revealed significant differences between the incompatible task and all other tasks.

Performance during saturation

First, all three tasks were analyzed for stability of performance during saturation. For tracking performance a repeated measures ANOVA was carried out with Task Difficulty (two levels: short Preview/large Amplitude (difficult) and long Preview/small Amplitude (easy)) and Session (four levels: Day 2 through Day 5) as factors. The intermediate difficulty levels could not be analyzed because the data of two subjects were not available. There was no significant variation in RMSE during saturation, $F(3,21) = 2.19$, $p = .119$. The poorer performance in the difficult task was still present, $F(1,7) = 108.49$, $p < .001$ and this difference remained constant across days, as shown by the absence of a significant interaction, $F(3,21) = .42$, $p = .741$.

For the choice-reaction-task (Focussed Attention task) a repeated-measures ANOVA showed that neither the factor Distractor condition (Neutral, Compatible, Incompatible, Single) nor the factor Session (with the four levels: day 2 through day 5) revealed a significant effect of Session on the average reaction time, $F(3,21) = 1.47$, $p = .25$, whereas the effect of Distractor was significant, $F(3,21) = 48.84$, $p < .001$. The effects of Distractor condition did not change across days as confirmed by a non-significant interaction between these factors, $F(9,63) = .85$, $p =.57$.

With the spatial-working-memory-task a repeated-measures ANOVA did not show a significant variation of distance error with time during saturation, $F(3,21) = .88$, $p= .446$.

It can thus be concluded that performance remained constant for all tasks across the four days of saturation.

Effects of hyperbaric condition

<u>Tracking task</u>. Figure 1 shows mean RMSE of the tracking task under normal and hyperbaric conditions. The comparison of performance between isobaric and hyperbaric conditions showed significant effects of tracking difficulty, $F(1,3) =$

103.80, $p < .002$, and pressure, $F(1,3) = 10.82$, $p < .05$, but no interaction, $F(1,3) = 1.65$, $p = .29$. Tracking performance became more variable; standard deviations showed an average increase of 12 % under 19 bar as compared to 1 bar of pressure, with this increase being more pronounced with the difficult tracking task (15 %) than with the easy one (10 %).

Table 1. Mean Root Mean Square Error, reaction time and distance error for the investigated conditions of the tracking, focussed attention and spatial memory task, respectively. Mean standard deviations are given in parentheses.

	Tracking (RMSE)			
Preview/Amplitude (ms/cm)	150-4.5	4500-4.5	150-1.75	4500-1.75
Baseline	3.114 (.287)	2.680 (.890)	1.836 (.214)	1.478 (.246)
Saturation	4.276 (1.086)	3.649 (1.371)	2.169 (.532)	2.123 (.590)
Difference	1.162	0.969	0.332	0.645
	Focussed Attention Task (RT in ms)			
	Single	Neutral	Compatible	Incompatible
Baseline	456 (36)	454 (32)	460 (41)	524 (28)
Saturation	533 (63)	548 (67)	539 (68)	585 (62)
Difference	77	94	79	61
	Spatial memory (error in mm)			
Baseline	3.350 (1.491)			
Saturation	4.281 (2.433)			
Difference	0.931			

Focussed Attention task. A two way ANOVA (4 distractor levels x 2 pressure levels) showed significant main effects of Distractor condition, $F(3,21) = 69.38$, $p < .001$, and hyperbaric condition, $F(1,7) = 62.55$, $p < .001$. The deleterious effect of hyperbaric condition on RT is in Figure 2. The negative effect was not the same in all distractor conditions, as shown by a significant interaction between the factors distractor condition and hyperbaric condition, $F(3,21) = 3.54$, $p=.032$. Post-hoc t-tests revealed no significant diffferences for the compatible, incompatible and single conditions ($p > .20$)

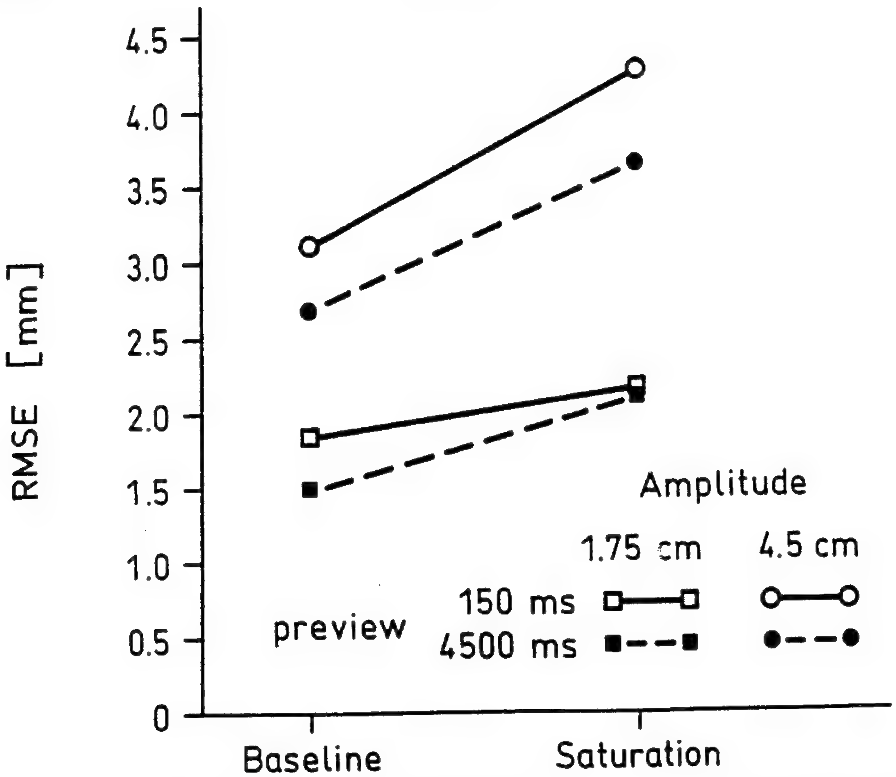


Figure 1. Averaged RMSE in baseline and saturation condition for the four tracking conditions

Spatial memory task. For the spatial-working-memory-task a t-test revealed a significant increase in the average distance error between performance in baseline condition and saturation, $t(3) = 2.47$, $p = .019$. Again, variability of performance increased by 12% during saturation.

Performance changes during the decompression phase

In Figure 3a-c the time course of performance across the seven days is shown for each task. A descriptive analysis of the performance on the three tasks during returning to normal pressure condition showed that in the tracking task the isobaric baseline performance of the easy condition was reached at 138 m, whereas in the other tracking conditions the improvement was more gradual. The spatial memory task also showed a rapid return to baseline. Here the performance at 138 m (3.6122 mm) was about the same as baseline performance (3.350 mm). A gradual return to baseline performance was found for all distractor conditions in the focussed attention task.

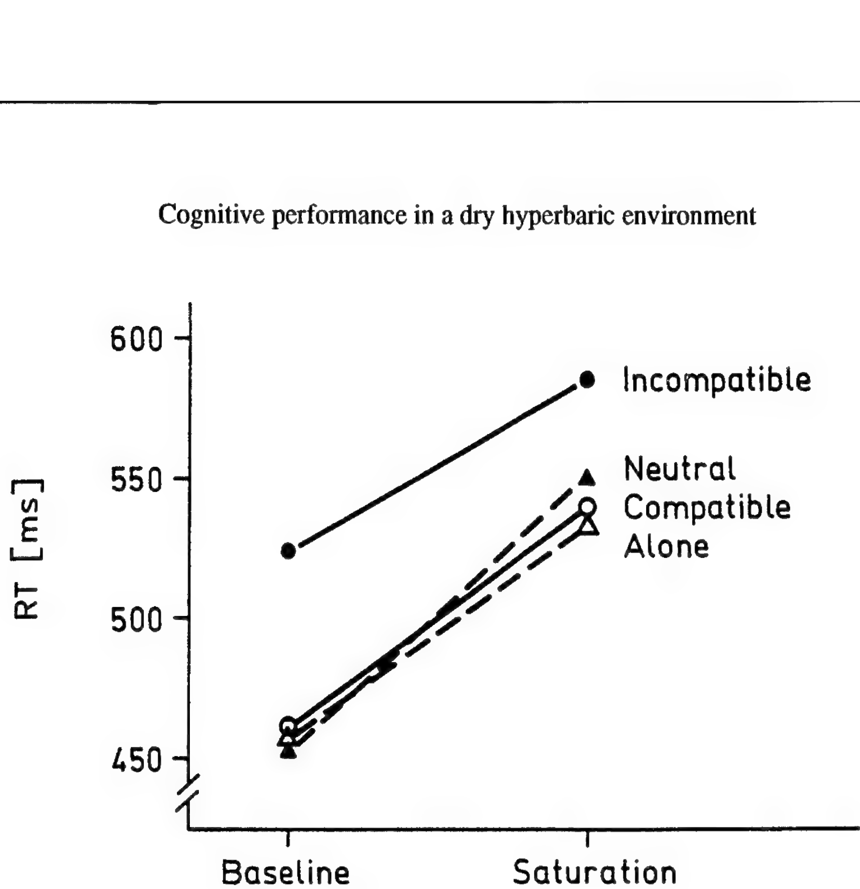


Figure 2. Mean reaction time in baseline and saturation conditions for the four conditions of the Focussed Attention task

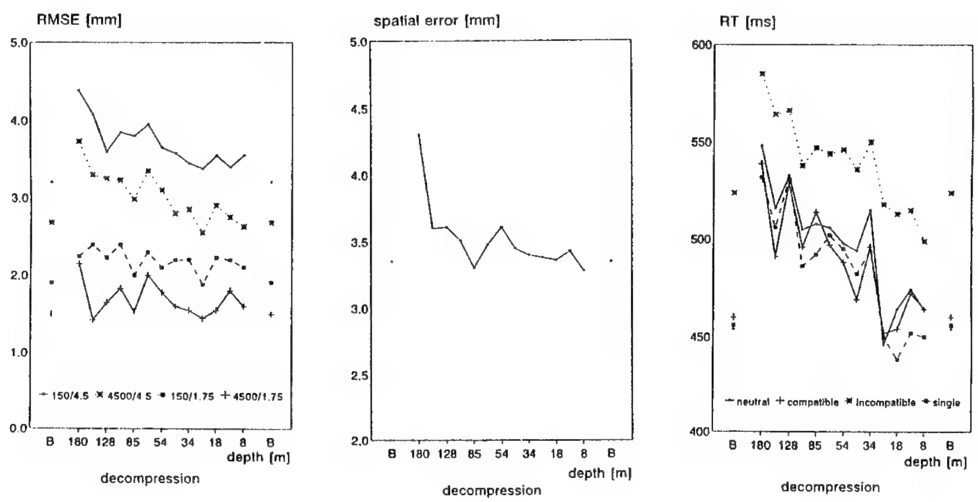


Figure 3a-c. Performance recovering during the decompression phase. From left to right the tracking task, the spatial memory task and the focussed attention task

Correlational Analysis

In order to investigate the relationship between performance in the three tasks under normal pressure (1 bar) a correlational analysis (Pearson r) was calculated.

Correlation values of the saturation condition are given in square brackets. The tracking task showed high correlations between the different difficulty levels (averaged r = .83) [r = .86]. A similar picture was observed for the four distractor conditions of the Focussed Attention task (r = .75) [r = .90].

Relatively high correlations were found between tracking performance and choice-reaction-time (r =.73) [r = .66] and between tracking performance and the spatial-working-memory task (r = .75) [r = .51]. However, the correlation between the choice-reaction task and the spatial working-memory task was low (r = .35) [r = .49].

Comparison between normal and hyperbaric conditions

Tracking performance under 19 bar showed a relatively high correlation with tracking performance under normal pressure (r = .76), with the average correlation being higher for the difficult task (r = .83) as compared to the easy one (r = .69). The average correlation between hyperbaric tracking performance with the baseline scores of choice-reaction-time was r = . 69. Again the correlation was higher for the difficult tracking task (r = .72) than for the easiest one (r = .67). The baseline scores of the spatial-working-memory task were somewhat lower than for the reaction task and also revealed a higher correlation with the difficult task (r = .63) than with the easiest task (r = .43) under hyperbaric tracking conditions .

Multiple Regressional Analysis for prediction of tracking performance under 19 bar by baseline performance scores did not show further statistically significant increase in predictive power as compared to correlation with the baseline correlations.

Discussion

Pronounced performance decrements were obtained by increasing the pressure to 19 bar in the hyperbaric condition. Both the criterion task of tracking (–40%) and its basic constituents –processing speed as measured in the Focussed Attention Task (–17%) and spatial working memory as indexed by the spatial memory task (–28%)– showed large performance decrements in the saturation condition as compared to baseline. Since performance did not show recovery during the saturation phase this performance decrease is a genuien result of the increased environmental pressure and not merely an effect of physiological adaption process to pressure during compression. The amount of decrement appeared to be independent of task difficulty. The notion of an additive effect of pressure on information processing is also supported by the fact that the difference in RMSE between the two tracking difficulties remained constant, irrespective of environmental pressure. At first glance this suggests that high pressure leads to a general slowing of information processing. However, while in the focussed attention task reaction time was prolonged, the ability to filter out response-

irrelevant signals was not affected differently. This means that high pressure does not cause a general slowing down of processing speed.

The correlational analyses show a clear relation between performance in tracking both for the reaction-time task ($r = .73$) and processing resources of spatial working memory ($r = .75$) under normal pressure. This supports the basic assumption of the Optimal Control Model that tracking performance is dependent on the one hand on the speed of information processing and on the other hand on the processing capacity of spatial working memory. At the same time, the relation between information processing time and capacity of spatial working memory is relatively low ($r = .35$) which indicates that the two components represent in fact two largely independent and qualitatively different basic processing factors of tracking. The fact that the observed correlations were similar indicates that both components play an equivalent role in tracking under normal environmental pressure conditions.

One factor that might be responsible for performance decrease in tracking could be the sedation caused by the narcotic effect of the gas mixture. The sedation could in turn retard speed of response execution. This is supported by the higher average correlation between tracking and the choice reaction time task in saturation ($r = .66$) as compared to the corresponding correlation between tracking and the spatial-working-memory-task ($r = .51$): both in tracking and in the choice-reaction-task processing speed is relevant for response execution, yet for the spatial-working-memory task this is not the case. Again this suggests that the effect of pressure is not general, but depends on the processes involved in the task.

Considering the fact that correlations between tracking and both reaction time and the spatial-working-memory task in saturation are lower than under baseline conditions, there seem to be still other factors responsible for the strong performance decrement in tracking. One of those factors could be response-execution inhibition caused by increased muscle tremor (HPNS symptoms). Similarly the increase of performance variability that was observed for all tasks under hyperbaric conditions could also be considered as reflecting instability of task performance induced by the HPNS syndrome.

However, because the reduction of the correlation under hyperbaric conditions is higher for the spatial memory task than for the choice reaction task, this might be interpreted as evidence that the tracking task decrement is caused by a decrease in the speed of processing of visual information. The slowing appears not to be related to problems of filtering out the relevant signal from the distractors, because the hyperbaric effect was additive to that of the distractor conditions. Thus, focussed attention processes seem not to be affected by increased pressure.

Most of all, the individual speed of processing of the tracking signal that is relevant for the task seems to be decisive for the inter-individual differences in tracking performance. This is indicated by the weaker correlation between tracking in saturation and the baseline scores of the spatial-working-memory task as

compared to the high correlation with the baseline scores of the choice-reaction task. The correlation of the compatible condition of the Focussed attention task with tracking was even higher ($r > .90$) than that of tracking performance in saturation and tracking performance on surface, thus providing even better predictive power for the criterion task of tracking than the tracking-task itself offers.

Thus, the results support the possibility of selecting divers for optimal execution of certain tasks in dry hyperbaric conditions during the planning phase of a dive for efficiency-maximization. At the same time hints are given for a practical use of the concept of a theory-based predictor-selection for real tasks by presenting appropriate predictors and possible methods of operationalization. However, before using simulation tasks these results should be validated with regard to the real tasks, and their generalizability to different levels of pressure should be investigated.

References

Adam, J. J. , Ketelaars, M. , & Kingma, H., & and Hoek, TH. (1993). On the time course and accuracy of spatial localization: Basic data and a two-process Model. *Acta psychologica, 84*, 135 - 159.

Biersner, R.J., and Cameron, B.J. (1970). Cognitive peformance during a 1000 ft helium dive. *Aerospace Medicine, 41*, 918 - 920.

Eriksen, B.A. and Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143 - 149.

Eriksen, C. W., and Schulz, D.W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics, 25*, 249 - 263.

Fitts, P. M., (1954). The information capacity of the human motor system in controlling the amplitude of the movement. *Journal of Experimental Psychology, 47*, 381-391.

Fröhlich, J. (1993). *Vorhersage des Spurhaltens beim Fahren von Kraftfahrzeugen durch Trackingtests.* Doctoral dissertation. RWTH Aachen.

Gill, R., Wickens, C. D., Donchin, E., and Reid, R., (1982). The internal model: A means of analyzing manual control in dynamic systems. In *Proceedings of IEEE Conference on Systems, Man and Cybernetics.* New York: Institute of Electrical and Electronics Engineers, Inc.

Jorna, P. G. A. M., & Moraal, J. (1988). *The use of simulation in predicting operational performance (A new means applied to old problems).* Institute for Perception TN0, Soesterberg, The Netherlands.

Lewis, V. J., and Baddeley, A.D. (1981). Cognitive performance, sleep quality and mood during deep oxyhelium diving. *Ergonomics, 24*, 773 - 793.

Logie, R.H., and Baddeley, A.D., (1985). Cognitive performance during simulated deep-sea diving. *Ergonomics, 28*, 731 - 746.

McRuer, D. T., & Jex, H. R. (1967). A review of quasi-linear pilot models. *IEEE Transactions on Human Factors in Electronics, 8*, 231.

O'Reilly, J.P., (1974). Performance decrements under hyperbaric HeO_2. *Undersea Biomedical research, 1*, 353 - 361.

O'Reilly, J.P., (1977). Hana Kai II: a 17-day dry saturation dive at 18.6 ATA. VI: Cognitive performance, reaction time, and personality changes. *Undersea Biomedical Research, 4*, 297 - 305.

Reid, D. &. D., N. (1972). A pilot model for tracking with preview. In *Proceedings, 8th Annual Conference on Manual Control* (Wright Patterson AFB Ohio Flight Dynamics Laboratory Technical Report AFFDL-TR-72, 92). Washington, D.C.: U.S. Government Printing Office.

Ross, H. E. (1989). Perceptual and motor skills of divers under water. *International Reviews of Ergonomics, 2*, 155 -181.

Schilling, C. W., Werts, M.F., and Schandelmeier, N.R. (Eds.) (1976), *The Underwater Handbook: A Guide to Physiology and Performance for the Engineer*. New York: Wiley.

Wickens, C. D. (1984). *Engineering Psychology and Human Performance*. Columbus, Ohio: Charles Merrill.

Zinkowski, N.B. (1978). *Commercial Oilfield Diving (2nd Ed.)* Cambridge, USA: Cornell Maritime Press.

Negative brain potentials predict performance in reaction tasks

M. Falkenstein, J. Hohnsbein & J. Hoormann
Institut für Arbeitsphysiologie an der Universität Dortmund, Germany

Abstract

Event-related brain potentials (ERPs) are a potential tool for the analysis of cognitive processes and performance deficits in human factors research. We earlier identified two subcomponents of the P300 complex in 2-way choice tasks, called P-SR and P-CR, which are related to stimulus evaluation and response selection, respectively. In addition, the slow brain potential preceding the stimuli (SPN) is assumed to reflect preparatory processes. Large differences in SPN amplitude were observed, which depended on the error rate in choice tasks: subjects with few errors (GOOD) had a large SPN, while subjects with many errors (POOR) had virtually no SPN. Moreover the P-CR of POOR subjects was much smaller, and delayed in comparison with GOOD subjects, regardless of response latency, which was similar for both groups. It is concluded that POOR subjects did not sufficiently prepare for the task (small SPN), which delayed and weakened their response selection (late and small P-CR), thus causing their higher error rate.

Introduction

Cognitive processes are accompanied by the mass activity of specific brain areas, particularly the cortex. Such activity can even be measured on the scalp as phasic or tonic electrical changes, the event-related potential (ERP). Hence it appears feasible to use the ERP as a direct physiological measure of cognitive processes during information processing.

ERPs have several advantages over other physiological measures:

1) they are noninvasive;

2) they do not interfere with the task;

3) they can measure different cognitive processing stages in real time;

4) they can measure the dynamics of those processes in time;

5) they can potentially yield information about the effort or expenditure of resources for a given task, i.e. elucidate the ratio of effort to outcome.

An ERP consists of different components that are separated in time. The single ERP components usually have distinct maxima on the scalp (Fig.1) that are spatially and temporally separable.

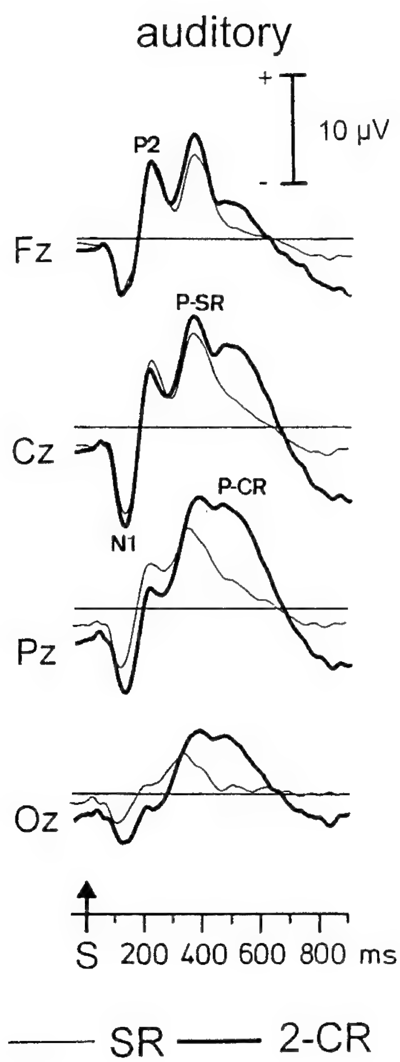


Fig.1. Example for an ERP following an auditory letter stimulus which had to be responded to by a simple reaction (SR) or by a two-way choice reaction (CR)(own data). Fz, Cz, Pz and Oz are different electrodes on the scalp (cf. Fig.4). The individual components have their maxima at different electrode locations (e.g., the N1 at Cz, the P2 at Fz, and the P-CR at Pz).

Usually the ERP follows a stimulus. However, there are also ERP components that occur before a task-relevant stimulus. One example is the so-called contingent negative variation (CNV), which builds up as a relatively slow negative deflection culminating at about the time when a task-relevant stimulus is expected.

The late part of the CNV, also called stimulus-preceding negativity (SPN; Brunia 1988; Rösler 1991) is also observed in a continuous series of task-relevant stimuli culminating at about the time of stimulus presentation (Fig.2).

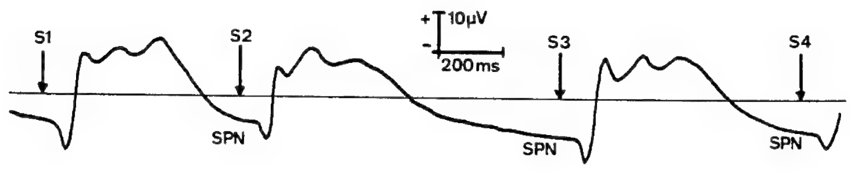


Fig.2. Semi-schematic example for a large stimulus-preceding negativity (SPN) at Pz developing before the stimuli (Si) in a continuous-performance two-alternative choice task. The horizontal line denotes technical zero.

The SPN is generally assumed to reflect, in psychological terms, the anticipation or preparation of the next trial, or in physiological terms, the facilitation of specific brain areas, which are relevant for the next trial (Brunia 1988; Rösler 1991).

One of the basic assumptions of ERP research is that the individual late ERP components reflect specific cognitive processes. The latency of a component reflects the timing of the process, while its topography gives information about the involvement of different brain areas in the process, and its amplitude is assumed to reflect the intensity of the process. It is crucial in ERP research to establish relationships between processes and components. With such relationships, variation in the amplitude and latency of the components can be used to infer, for example, the influence of specific work conditions on specific processing stages, or to specify the reasons for performance deficits. Our approach to establishing component-process relations is the observation of changes of ERP waveshapes under well-defined changes of task conditions.

During the past few years we focused on the largest ERP component, the P300. We found that the P300 comprises two subcomponents (Hohnsbein et al. 1991; Falkenstein et al. 1993). The first has a different latency and topography for visual and auditory stimuli and is associated with stimulus evaluation (identification). Since it is also present in simple reaction tasks, we called it P-SR. The second positivity has a clear parietal maximum regardless of stimulus modality. Since it is present only in choice reaction tasks, we called it P-CR. Examples of these components are shown in Fig.1. When response selection complexity is manipulated, the P-SR remains constant in latency, while the latency of the P-CR is strongly influenced (Falkenstein et al. 1994). Moreover the P-CR is larger for more complex than for easy choice tasks. This lead us to the conclusion that the P- CR is related to the response selection process. The amplitude of the P-CR appears to reflect the complexity of, or in other terms, the resources allocated

to, the response selection process, and the P-CR-latency is related to the timing of response selection.

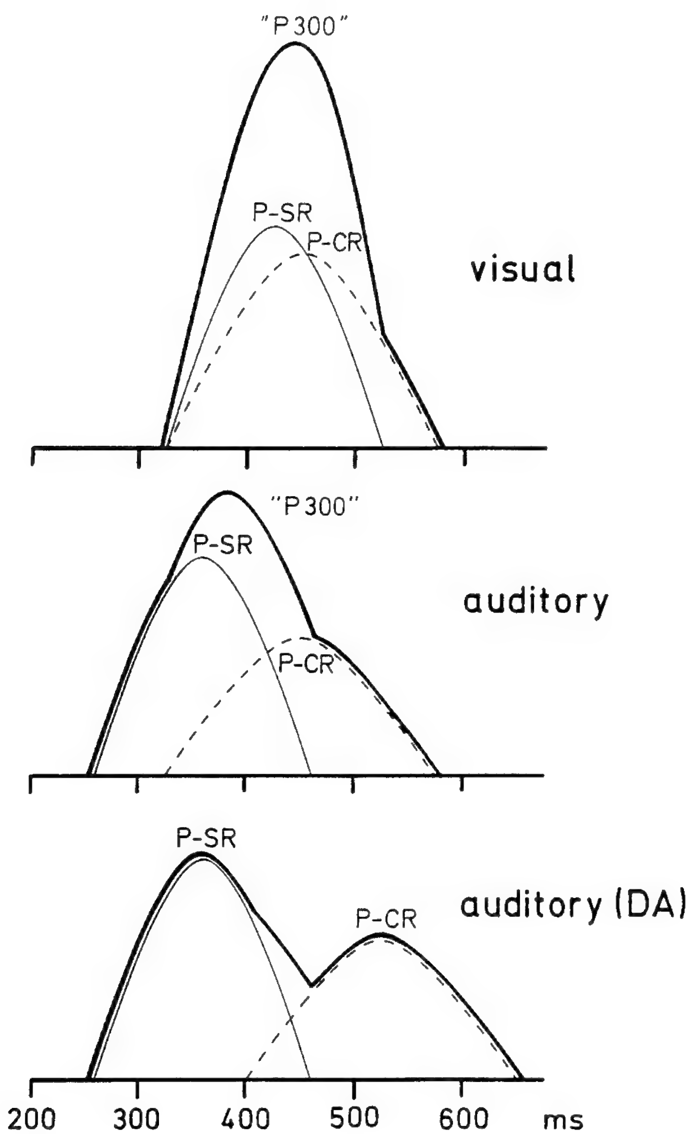


Fig.3. Model of two subcomponents, P-SR and P-CR, of the P300 complex in two-alternative choice tasks. The subcomponents merge to one late positive complex "P300" after visual stimuli (upper panel), whereas the separation of P-SR and P-CR is better after auditory stimuli, where the P-SR peaks early (middle panel). The best separation is achieved after auditory stimuli, when the stimulus modalities are mixed within a block (DA (divided attention) paradigm, which is used in the present study; lower panel).

Thus the P-SR and the P-CR are tools that may shed light on the timing and the intensity of brain processes that are linked to stimulus evaluation and response selection, respectively. Fig.3 shows the two components and their summation to the P300 complex. The disentangling of P-SR and P-CR is more complete for auditory stimuli, since for these the P-SR occurs earlier than for visual stimuli. An even better separation of the subcomponents is achieved for auditory stimuli if they are intermixed with visual stimuli within a block (divided attention (DA) paradigm; Hohnsbein et al. 1991).

The present study was designed to investigate whether subjects with different performance accuracy, as defined by their error rates in a choice task, also differ in the structure of their event-related potentials (ERPs). The divided attention paradigm was used to obtain a better separation of the P300 subcomponents. Our hope was to find reasons for the particularly large difference in performance (error rate) found among the subjects in a choice reaction experiment. A potential tool for predicting performance is the SPN or late CNV, since it appears to reflect preparation, as pointed out earlier. In studying the ERPs following the stimulus we focused our interest on the subcomponents of the P300 complex, P-SR and P-CR, in order to infer whether stimulus evaluation or response selection, or both, were conducted in a different way in good and poor performers.

Methods

Ten highly-trained subjects performed two-way choice reactions and simple reactions to visual or auditory letters (F and J). The visual letters (0.5 deg high) were presented for 200 ms in the middle of the screen of a visual display unit (VDU). The (spoken) auditory letters were stored in the RAM of a micro-computer and presented diotically via headphones. Letters and stimulus modalities were randomized within a block, which contained 50 stimuli of each type (auditory F, auditory J, visual F, visual J). The interstimulus interval was randomized around 1500 ms (between 900 and 2100 ms). A moderate time pressure was imposed by a feedback signal on trials that gave reaction times beyond a fixed limit (350 ms for simple reactions, and 500 ms for choice reactions). Each block was presented twice. The EEG was recorded from the midline electrodes of the 10-20 system (Jasper 1958), i.e., Pz, Oz, and additionally from the two central-lateral electrodes C3 and C4. The EEG was amplified 100,000x and band-pass-filtered 0.03 to 60 Hz. ERPs of correct and incorrect trials were averaged separately, using both the stimulus and the response as trigger. In the averaged ERPs the mean of the 50-ms period before stimulus onset was defined as the SPN; for the ERP components following the stimulus the peak latencies and amplitudes were evaluated. These parameters were evaluated statistically by analysis of variance. For the correct trials the factors were group (GOOD, POOR) and the repeated measures factors electrode (Fz, Cz, Pz, Oz) and stimulus modality (auditory, visual). For the reaction times (RT) the electrode factor was of course omitted.

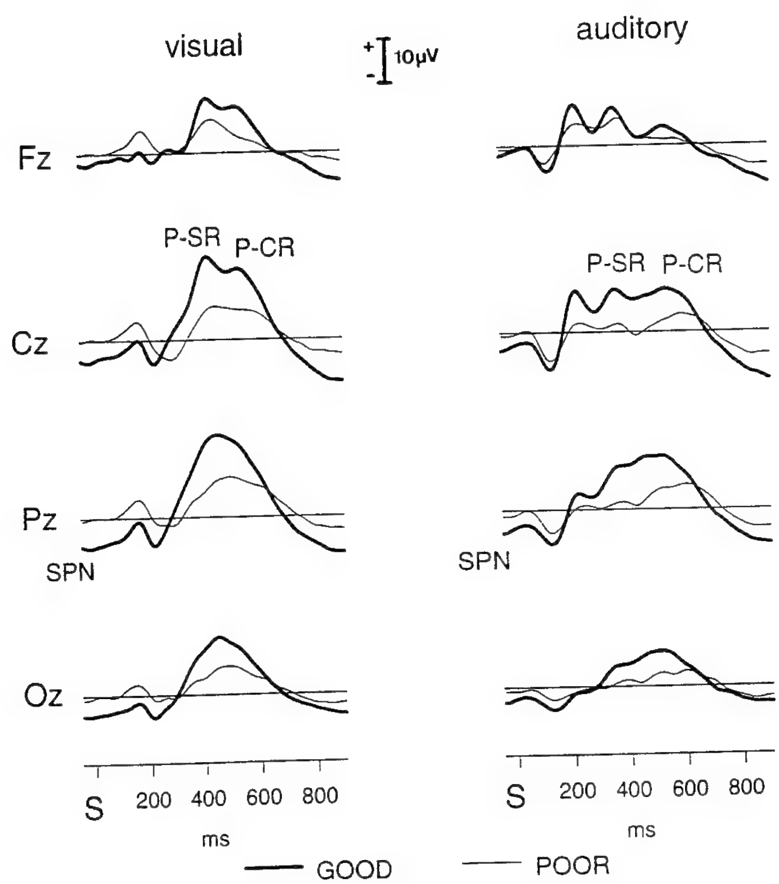


Fig.4. Grand averages (mean waveshapes of the ERPs across subjects) of the ERPs after visual (left) and auditory stimuli (right) for GOOD subjects (heavy lines) and POOR subjects (thin lines). The SPN (the negative displacement before stimulus onset) is nearly absent for POOR subjects. The P-CR is smaller and delayed for POOR subjects, which can be best seen for auditory stimuli.

Results

Due to the time pressure the average error rate in the choice task was 13%. The distribution of error rates was bimodal: Five Ss (termed "GOOD") had a relatively low error rate (about 6%), while it was high (about 20%) in the other five Ss (termed "POOR"). GOOD and POOR Ss were not pre-selected, but all 10 Ss participating were grouped according to their performance in this experiment. Altogether the RTs of correct trials were about 370 ms. Generally GOOD

performers had slightly longer RTs (377 ms) than POOR performers (367 ms). However, this difference was only significant after auditory stimuli (p=.03).

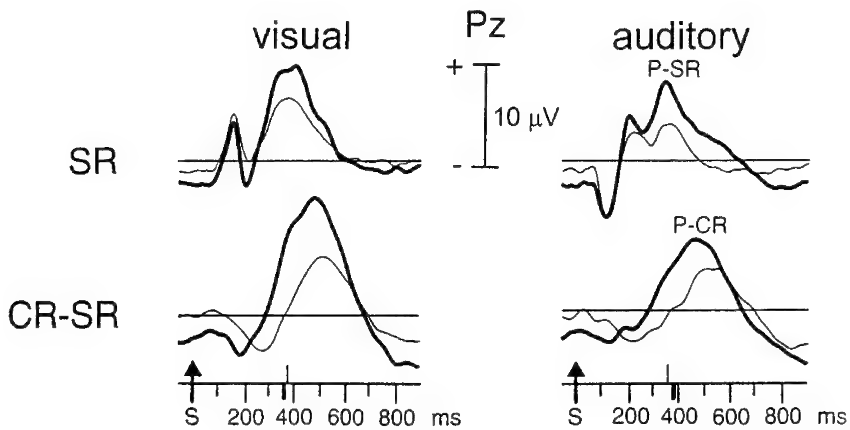


Fig.5. Grand averages of the ERPs of the simple reactions (SR) and of the difference waves between the ERPs of choice and simple reactions (CR-SR) (Pz electrode) for GOOD subjects (heavy lines) and POOR subjects (thin lines). The difference waves suppress early components and P-SR and highlight those components that only occur in the choice task. The amplitude and latency differences are not only seen for the P-CR, but also for the negative wave preceding the P-CR.

The ERPs of correct choice-reactions (Fig.4) revealed a large stimulus-preceding negativity (SPN) with parietal maximum (about −5 mV) for GOOD performers, and only a very small SPN (about −0.5 mV) for POOR performers. This group difference was highly significant (p = .002). The early ERP components (the visual P170 and the auditory N140) showed no significant amplitude or latency difference between GOOD and POOR performers (although it may appear so from the grand means). Both late positive components appear to be larger for GOOD than for POOR performers. In fact, the group effect on amplitude was rather weak for the P-SR (p = .02) and strong for the P-CR (p = .002). A closer topographical analysis showed that the group differences were largest at Pz for the P-CR (which had in fact a Pz maximum) as well as for the P-SR (which had a fronto-central maximum) which suggests that the apparent P-SR-effect was due to the strong enhancement of the overlapping P-CR for GOOD compared to POOR performers.

In contrast to the RTs the latency of the P-CR was about 50 ms longer for POOR than for GOOD performers (p = .001), whereas the P-SR latency was the same for both groups. This is best seen in Fig.4 for the auditory ERPs, where the separation of P-SR and P- CR is larger than for visual ERPs.

In order to suppress the early ERP components and the P-SR, which also occur in simple reaction ERPs, we subtracted the latter from the choice reaction ERPs. The difference waveshapes highlight those ERP components that are restricted to choice tasks. They clearly confirm the P-CR results with respect to latency amd amplitude and show in addition a negative wave before the P-CR, called N-CR, which has a similar latency shift across groups (Fig.5).

The main findings can hence be summarized as follows: The SPN was larger, and the P-CR was larger and earlier for GOOD than for POOR Ss. The earlier components, including the P-SR, showed no group effects.

Error trials

In the error trials the RTs were about 14 ms shorter (358 ms) than in correct trials (372 ms). The SPN showed a similar, though somewhat smaller, difference between GOOD (–3.2 mV) and POOR (–0.3 mV) performers (p = .03) than in the correct trials.

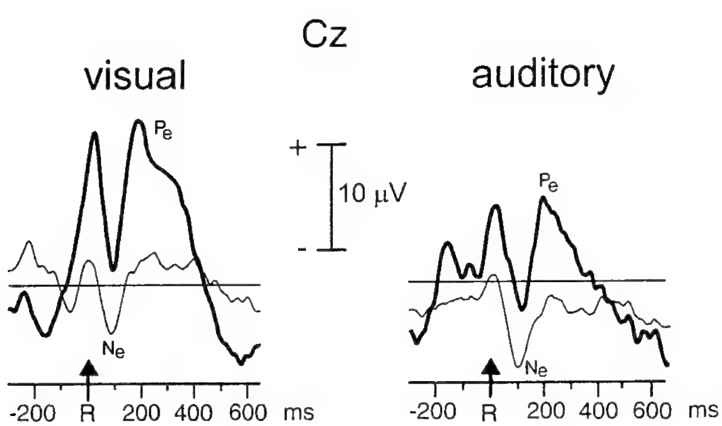


Fig.6. ERPs of error trials, averaged with the response (R) as trigger. (Cz electrode, where both error-related components have their maximum.) The error negativity (Ne) is similar in both groups, while the error positivity (Pe) is virtually absent in POOR subjects.

Earlier we showed that on error trials a negative component (error negativity, Ne) and a subsequent positive component (Pe) are elicited, which most probably reflect different aspects of error processing (Ne: error detection, Pe: further controlled error processing.) These components are best seen in the response-locked averages at Cz (Fig.6).

In the present data the error negativity (Ne) peaked at about 60 ms after the incorrect key press. There was no significant difference in Ne amplitude or latency between GOOD and POOR subjects. In contrast, the Pe was large for GOOD and virtually absent for POOR performers ($p = .002$).

Discussion

The results show that a strong preparation, as reflected in the SPN, is associated with a large amplitude, particularly of the ERP component (the P-CR) that is related to the main controlled process, response selection. Moreover subjects with large SPN (GOOD) began and finished the cognitive response-selection process (as reflected in P-CR latency) earlier than subjects with small SPN (POOR), independent of the choice RT, which tended to be somewhat longer for GOOD subjects. Similar results with respect to the amplitude (but not the latency) of a late positivity similar to the P-CR, and to an SPN development only in GOOD Ss, have been found by Brookhuis et al. (1983), and were interpreted as a sign of reduced processing capacity in POOR Ss. This can also be concluded from our data, with the additional claim that this reduced capacity is due to a lack of preparation, as reflected in the SPN. The intense and fast cognitive response selection, and the tendency to withhold the response until sufficient evidence from response selection is available, can explain the low error rate for GOOD subjects. In contrast, POOR subjects combined weak and slow cognitive response selection with a tendency toward fast (premature) responses, which can explain their high error rate. The results further indicate that stimulus processing is most probably not different among the groups. It may be argued that late positive potentials of the preceding trial affect the development of the SPN before the following one. However, because P-CR and Pe are very small for POOR subjects, the possibility that they could prevent a sufficient negative shift before the next trial is unlikely. On the other hand, GOOD subjects build up a large SPN despite the fact that it is preceded by a large P-CR. The finding of faster reaction times and similar SPN on error compared to correct trials shows that the actual errors were made because of particularly fast premature responses. The similarity of the Ne across performance groups showed that errors were detected equally well in both groups. In contrast, the absence of the Pe in POOR subjects shows that the further controlled error processing is strongly reduced in these subjects. This may mean that POOR subjects regard frequently-occurring errors as unimportant.

Conclusion

The amount of preparation, as reflected in the amplitude of the stimulus preceding negativity (SPN), strongly influenced the timing of, and the resources allocated to the cognitive process of response selection, as reflected in the P-CR. In particular, a small SPN was associated with a delay of response selection. Given the need for a fairly fast response (time pressure) the high error rate in poor subjects is caused by their slower cognitive processing and their tendency to give premature responses.

References

Brookhuis, K.A., Mulder, G., Mulder, L.J.M., Gloerich, A.B.M., van Dellen, H.J., van der Meere, J.J., and Ellermann, H. (1983). Late positive components and stimulus evaluation time. *Biological Psychology, 13*, 107 - 123.

Brunia, C.H.M. (1988). Movement and stimulus preceding negativity. *Biological Psychology, 26*, 165 - 178.

Falkenstein, M., Hohnsbein, J., Hoormann, J. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and clinical Neurophysiology, 78*, 447 - 455.

Falkenstein, M., Hohnsbein, J., Hoormann, J. (1993) Late visual and auditory ERP components and choice reaction time. *Biological Psychology, 35*, 201 - 224.

Falkenstein, M., Hohnsbein, J., Hoormann, J. (1994). Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography and clinical Neurophysiology, 92*, 148 - 160.

Hohnsbein, J., Falkenstein, M., & Hoormann, J. (1991). Effects of crossmodal divided attention on late ERP components. I. Simple and choice reaction tasks. *Electroencephalography and clinical Neurophysiology, 78*, 438 - 446.

Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and clinical Neurophysiology, 10*, 370 - 375.

Rösler, F. (1991). Perception or action: some comments on preparatory negative potentials. In C.H.M.Brunia, G.Mulder and N.Verbaten (Eds.) *Event-Related Brain Research* (pp. 116 - 129). Amsterdam: Elsevier.

3-D Radar Display for Air-Traffic Control Tasks

M. Göbel, J. Stallkamp & J. Springer
Institute of Industrial Engineering and Ergonomics,
Aachen University of Technology,
Germany

Abstract

Air-traffic control tasks are based on an overview of the actual air traffic situation. The major part of this information is transmitted visually by a radar screen. In conventional screen designs the air traffic is displayed from a bird's eye-view in a two-dimensional form. Since the actual areas to be controlled are three-dimensional, the missing dimension on the screen, the altitude, is printed on numeric labels beside the aircraft. Through observation of the radar screen the air-traffic-controller builds a mental model of the actual traffic situation. For this reason, air-traffic-controllers have to spend a considerable part of their mental capacity in transforming the radar screen information. This requires long experience and causes an increase of workload as limiting their tracking capacity. The present paper describes a prototype three-dimensional display for air traffic control.

Introduction

This paper describes the development of a new three-dimensional radar screen. The aim of this new design is to transmit spatial information in a direct way to the air-traffic controllers and thereby to reduce the human information processing demands. Using a CRT with an active polarisation filter in front of the screen and passive polarisation filters in front of the user's eyes, information is presented to the two eyes separately by switching the polarisation direction. The disparity of left and right eye images leads to the perception of object-positions in front of or behind the physical surface on the screen. This technology was used to create a 3D-perception comparable to a natural view. It was not our intent to simulate a natural view completely, but to use the additional dimension to display object interactions in space as they would appear in reality. Therefore a symbol-coded display with different scales in the horizontal and vertical direction was chosen. To enable a precise evaluation of specific situations, the viewpoint and the (virtual) viewing distance can be controlled in all 6 degrees of freedom by a space-ball.

During different presentations, air-traffic controllers expressed the opinion that this new design might lead to improved working conditions. For the future, a comparative evaluation of both conventional and 3-D displays is planned in order to quantify possible improvements of a three-dimensional representation on workload and task performance.

Air-traffic control tasks

In air-traffic control (ATC), the performance and the security of air traffic is largely determined by human operators. The work of air-traffic controllers is a highly complex task with numerous, highly varying, workload-factors. Besides the primary task of route planning, it is strongly associated with its organisational, social, and technical environment.

The whole airspace is divided into an upper area control (UAC, above 24000 ft.) and a lower area control (LAC, below 24000 ft.). In lower area control, particularly in the vicinity of an airport, air traffic density increases and requires frequent changes in altitude. To ensure guiding capacity for the whole airport, the airspace is divided into different sectors which are controlled separately by different controllers and which have to be coordinated.

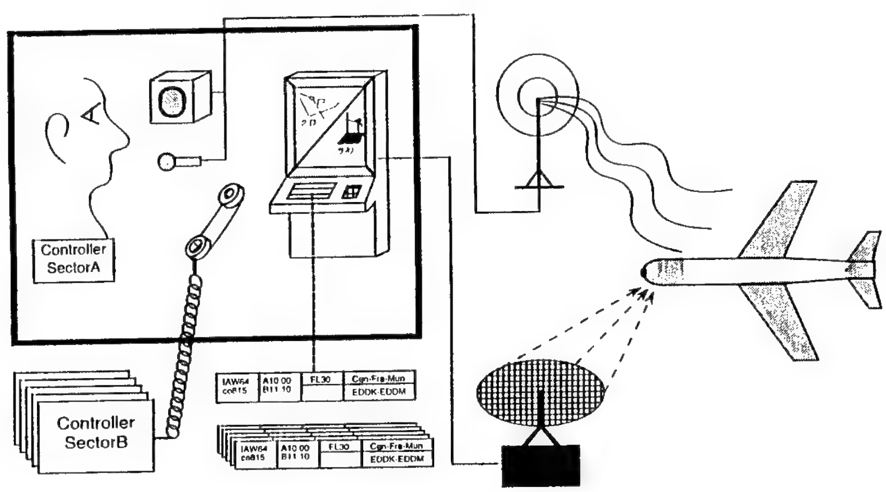


Fig. 1. Communication and interaction structure for air-traffic control tasks

The major information about the current air traffic situation is given by a radar screen, enhanced by the flight plan in form of flight-strips on paper or in electronic form. Output communication is achieved by a speech link to the pilots and by telephone to the controllers of other sectors. During take-off and landing aircraft have to be guided on glide paths (usually by vector-navigation) at intervals as short as approximately 60 seconds, requiring a precise timing of the whole air traffic. Nevertheless, no errors that could endanger lives can be tolerated.

Since air traffic controllers have to communicate exclusively by technical means, the radar screen is the main information source about the actual air traffic

situation*. In traditional form, the air traffic is displayed in a two-dimensional form from a bird's eye view, enhanced by a special map on the same screen (including glidepaths, sector limitations, restricted areas and radar-zones). The missing dimension on the screen, the aircraft altitude, is displayed in numerical form on labels beside each aircraft symbol. The aircraft information is further enhanced by the flight number (aircraft callsigns), an attitude indicator and a velocity vector. Further electronic information services have been developed during the last few decades, which lead to numerous screen designs. To improve usability, the latest generations of radar-screens combine all these information sources in a multi-window layout (including tele-communication window, 10-minute entry warning window, screen management menu, conflict risk display and a message window).

Workload of air-traffic controllers, bottlenecks of the technical environment

One of the most serious bottlenecks of such a highly-complex system is the information exchange between the technical system and the human operators. Since it is not possible to perceive all information about the actual air-traffic situation at once, the controller has to scan the radar display continuously for information of situational relevance and to build up a mental model in a first step. From this mental model, further decisions and actions are derived.

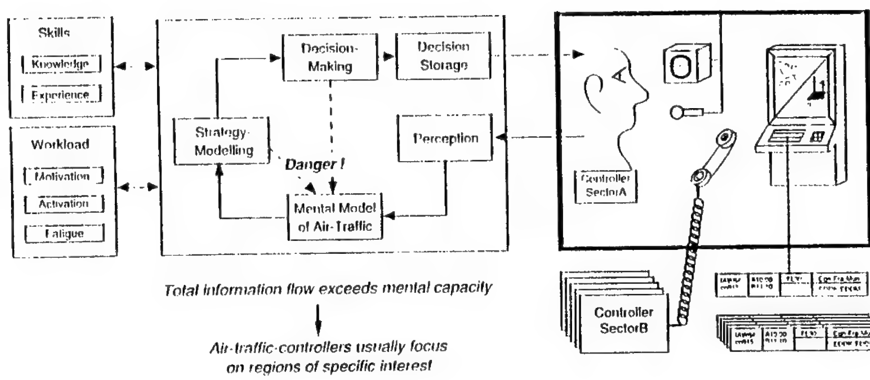


Fig. 2: Simplified information-processing structure for air-traffic control tasks.

This procedure implies some specific restrictions:

- Mental effort is partially wasted in the additional translation process needed to form the mental model, resulting in increased workload.

* Although air-traffic control must be assured in case of loss of visual information about the air traffic situation (failure of the radar system)

- Situations of high air traffic density might exceed the controller's mental capacity and therefore limit the tracking capacity[*].

- To keep an accurate model of the actual traffic situation at hand, intense concentration has to be maintained. Even short interruptions in observation can cause difficulties in continuing the task.

- Error probability is increased due to the exclusively mental representation of the spatial relations. Since for strategy planning a second (mental) model is required, undesired crosstalk between both mental models can cause ineffective routing and critical situations.

These high demands on the controllers' performance, in conjunction with an increasing in air traffic density, lead to specific consequences for work organisation and economic factors:

- Air-traffic controllers are required to be very experienced. An extensive course of training must be passed and a remarkable percentage of trainees fail the final tests (in Germany 75 to 90%; Eissfeld et al. 1993).

- Critical situations out of the focus of attention of the air-traffic controller might not be recognized.

- The high stress levels implies a relatively low retirement age in this profession. As early as 1973 Rohmert found that 75% of controllers did not expect to be able to continue working to the normal retirement age. In fact, about 50% of controllers were between 30 and 35 years old, and only 22% were older than 39 years.

In consequence, many efforts are made to improve the working conditions of air traffic controllers through the application of new technologies. Technologies improvements for air traffic control may start from two points:

- Shifting tasks, at least partially, from the human being to 'intelligent' technical support systems (e.g. collision avoidance systems, automatic data communication or automatic route control for standard procedures), and

- optimizing communication and interaction between the technical system and the human air-traffic controller.

Activities in these fields are not mutually exclusive but rather they rely on each other. Current tendencies to automate air-traffic control tasks will increasingly shift the controllers' task from active guiding to surveillance. In consequence, this requires more powerful information exchange between the technical system and the human operator, due to the increased tracking responsibility for each subject.

[*] In the context of air-traffic control, tracking capacity refers to the number of aircraft to be controlled.

Furthermore, any intervention in automated procedures requires an intensive bi-directional information exchange between man and machine.

With reference to the functional mechanisms of human information reception and processing a significant potential for design improvement can be expected, if the information representation is selected as similar as possible to the operator's mental structure of the natural environment. In this case, the controller would be able to make use of physiological attributes and life-long experience with spatial orientation. At this point it has to be noted that there is still a lack of knowledge about the structure of mental modelling during air traffic control tasks. First projects have been initiated dealing with the mental modelling of air-traffic controllers (e.g. the EnCoRoute-project of the DFG, Bierwagen 1993). Furthermore, a strong relationship between the mental model and the information that it represents has to be assumed.

Requirements for the visual representation of spatial relations

Due to the three-dimensional structure of air traffic (with time as a fourth dimension), a two-dimensional visualisation limits but not precludes spatial perception. Considering the cues for spatial image recognition (according to Wickens, 1992) a total of eight monocular and two binocular cues can be identified.

This ratio does not mean that the binocular cues are of less importance because they are less numerous, but each cue has a different meaning depending on the kind of image, viewing distances and lighting conditions.

An air traffic situation displayed in a perspective view on a two-dimensional screen (2.5-D) utilizes only the cues 'object overlap' and 'relative object size' as indicators for depth; the other cures cannot be displayed due to the short viewing distance and other technical restrictions (resolution, sharpness). In the mid-eighties some experiments with these kinds of displays were reported (e.g. Huffman & Verschaffel 1985, McGreevy & Ellis 1986). In most cases a 2.5-D system with a perspective presentation on a two-dimensional screen was used, but even with the additional application of indication lines to avoid misinterpretations the results were not satisfactory. The major restriction was the inability of the controllers to judge the aircraft positions unambiguously.

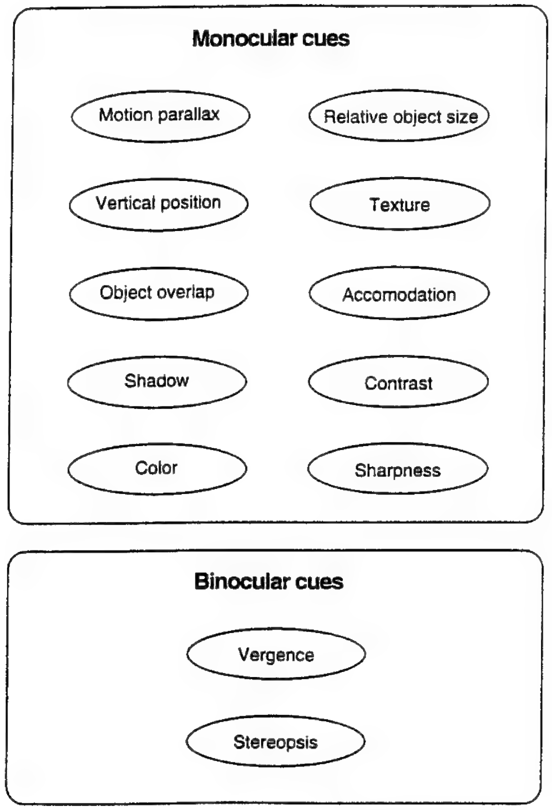


Fig. 3: Visual cues for human spatial perception

In consequence, the spatial representation of air-traffic-situations requires the use of the binocular cues 'vergence' and 'stereopsis' at least, if appropriate enhanced by monocular cues.

Realisation of a 3-D radar display

In order to evaluate a true three-dimensional space perception in ATC-applications, a demonstration model of a 3-D radar screen was developed.

To do this it is necessary to generate and to transmit two different images to both eyes. For a pre-determined viewing distance and interocular distance, the corresponding images can be calculated from the radar data so as to be similar to the retinal images of real objects. A common modern technique for presenting different images to the two eyes displays both images alternately on a CRT screen and then uses electronic shutter glasses to control the information flow to the eyes. A more awkward technology, using the same principle, can be achieved by mounting an active polarisation shutter in front of the screen and using passive stereoscopic colour images, but they require a doubling of the CRT repetition frequency compared to a monoscopic display. Polarizing shutters were preferred for

air-traffic-control applications since the passive polarisation glasses (which resemble light-toned sunglasses) minimize discomfort and do not interfere with other equipment.

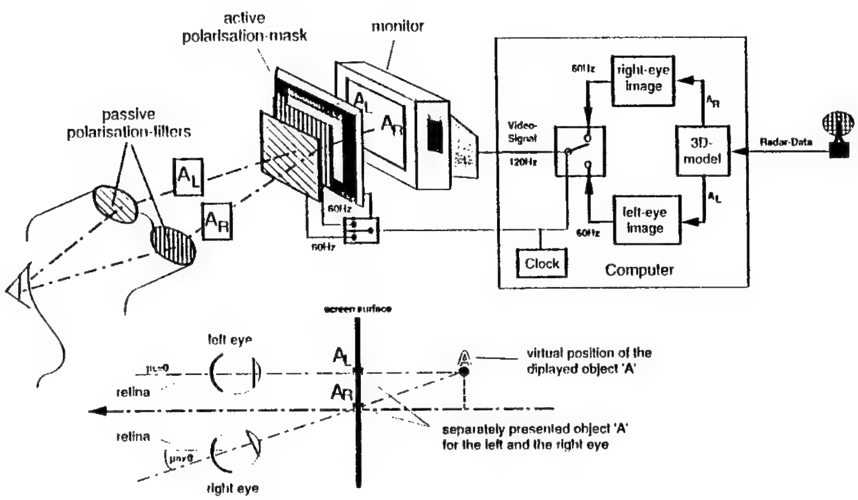


Fig. 4: Principle of stereographic image generation using polarisation shutter technology.

On the graphical side, the air-traffic situation is translated into a symbol-orientated screen layout, but with a correct spatial representation. Aircraft are displayed by a wedge indicating their actual position and direction. To enable the precise labelling of an aircraft's geographic position, which is obviously not possible because of the perspective view, the base surface is represented by a grid with fixed distances and a vertical bar between the aircraft and the base surface. Object movement and speed are further indicated by a line behind the aircraft, representing its past path during a fixed period of time. To assist precise guiding, localizers and glidepaths are additionally displayed. Object identification is shown as usual with labels in text or numeric form, whereby the three-dimensional presentation helps to avoid that labels are hidden by other objects. Additionally, all fixed objects (e.g. landing zones and zones of bad weather) are displayed in different colours. In regions with specific topography, such as mountains, valleys and sea-coasts, these objects are included in order to support appropriate guiding strategies.

To adapt to specific situations, a 6 degrees of freedom spaceball is connected to the graphic generator which allows the user to move the viewpoint and to change the viewing field corresponding to the ball-movement.

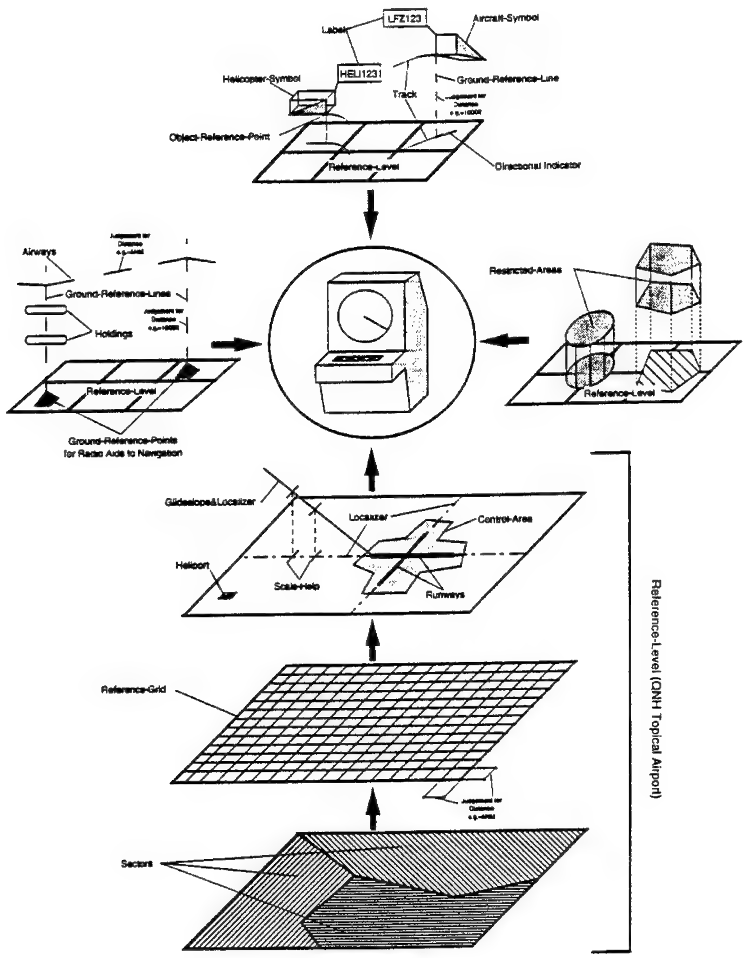


Fig. 5: Screen composition of a 3-D radar display.

The demonstration model leads to the expectation of some important benefits for air-traffic-controller's work:

- information density is reduced remarkably without losing or suppressing any information due to the spatial coding of aircraft positions,
- security can be increased, especially for aircraft out of the centre of the controller's field of view.
- increased tracking capacity, tracking precision, and guiding quality (e.g. shorter holding patterns, better coordination of different aircraft),
- spatial representation of restricted areas and zones of bad weather,

- no label overlapping,
- decreased training requirements,
- reduced workload, and a
- later retirement.

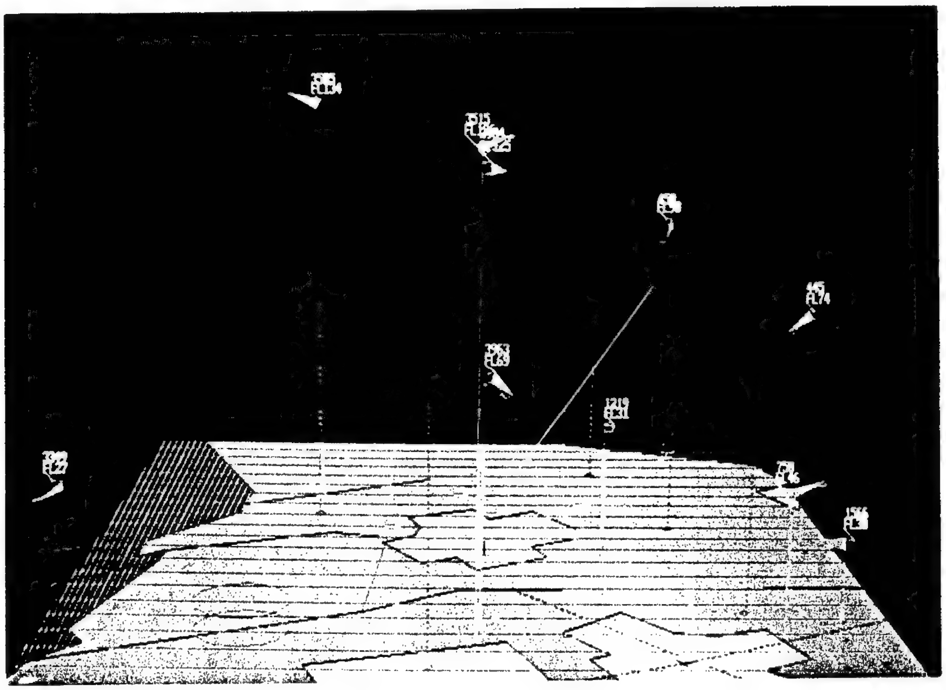


Fig. 6: Monoscopic print of the 3-D demonstration display.

On the other hand, specific implementation constraints have to be considered:

- fixing the exact coordinates of an aircraft is not yet possible in a direct way,
- air traffic controllers have a very high degree of experience working with conventional displays and therefore any change requires additional effort to adapt,
- long-term workload-factors of working with 3-D shutters are not yet known*, and

* The divergence between the vergence of the eyes, fixating a virtual object, and the accommodation to the physical screen surface might lead to additional visual fatigue.

- the implementation of shutter technologies with $\geq$ 120 Hz image frequency on a 2000 x 2000 pixel monitor (27") is currently not available.

Future perspectives

Although this model was generally judged favourably by experts and users, an objective evaluation of its influence on ATC performance and controller's workload has not been performed.

To facilitate the adaptation process for experienced air-traffic controllers, a switchable screen design for conventional 2-D and the new 3-D representation should be applied. Thus the controller might select the most appropriate mode of visualisation for himself. Furthermore, new input devices might be applied to complete the interaction according to the three-dimensional presentation, such as communication control by pointing at the selected aircraft of aircraft guiding by direct object manipulation (direct grasping and tracking of the aircraft).

In summary, a three-dimensional radar screen provides the potential to increase performance and to decrease workload compared to conventional screen layouts. Especially if future technologies to assist the guiding task are introduced, the increased information flow to the controllers will have to be managed. A 3-D design might be one step toward satisfying these demands. A critical factor for such a design is the complete change of information perception and probably of the working structure; a step-by-step evolution form the conventional screen layout does not seem possible: therefore the implementation of a 3-D radar screen has to be considered as a long-term project.

References

Bierwagen, T. (1993) En-Route Controller's Representation. DFG-Projekt *Flugsicherung als Mensch-Maschine-System*, Berlin: Technische Universität Berlin, Institut fur Psychologie.

Eissfeld, H., Goeters, K., Rathje, H. (1993). *Eignungsauswahl fur den gehobenen Fulgverkehrskontroll-und Flugdatenverarbeitungsdienst*. Deutsche Forschungs- und Versuchsanstalt fur Luft-und Raumfahrt e.V., interner Bericht IB 316-93-02.

Hopkin, D.V. (1982). *Human Factors in Air Traffic Control*. AGARD Report AG-No.275.

Huffman, A., Verschaffel, M. (1985). *A Dynamic Three Dimensional ATC View*. EUROCONTROL Experimental Center Task No. D22/4.G.

McGreevy, W., Ellis, S.R. (1986). The Effect of Perspective Geometry on Judged Direction in Spatial Information Instruments. *Human Factors, 28*, 439 - 456.

Rohmert, W. (1973). Psycho-physische Belastung und Beanspruchung von Fluglotsen. In Schriftenreihe *Arbeitsmedizin, Sozialmedizin, Präventivmedizin*. Stuttgart: Gentner Verlag.

Wickens, C.D. (1992). *Engineering Psychology and Human Performance (2nd Ed.)*. New York: HarperCollins.

Accommodation, convergence, pupil and eye blinks at a CRT-display flickering near fusion limit

Wolfgang Jaschinski, Matthias Bonacker & Ewald Alshuth
Institut für Arbeitsphysiologie an der Universität Dortmund
Germany

Introduction

Visual display units (VDU) with cathode ray tubes (CRT) refresh the CRT phosphor periodically at the frame frequency of the VDU. This can give rise to the perception of flicker and, consequently, visual discomfort and asthenopic complaints. Flicker disappears if the refresh rate exceeds a limit, called critical flicker frequency (CFF), which typically lies in the range of 50 - 100 Hz, depending primarily on the sensitivity of the subject, the particular viewing conditions, and the CRT phosphor decay time. The average critical flicker frequency for a bright-background CRT screen is around 70 Hz. Perception of flicker can be avoided by using refresh rates above the CFF.

However, absence of visible flicker does not necessarily mean that all visual functions have reached a state corresponding to steady light. Several studies have uncovered evidence that visual functions may respond to intermittency of light even if the refresh rate exceeds the critical flicker frequency. The electroretinogram (ERG) responds to frequencies above CFF. Berman et al. (1991) described synchronous ERG-responses to a special text arrangement on a CRT screen with a refresh rate as high as 76 Hz (which was above CFF for these stimulus conditions) and ERG-responses up to 145 Hz elicted by directly viewed fluorescent lamps. On the other hand, visually evoked cortical potentials have a cut-off frequency near CFF (Sternheim and Cavonius, 1972).

Other studies have investigated the focusing mechanism of the eye and measured the static accommodative responses to flickering monocular stimuli as a function of the viewing distance. For near targets, the accommodative response (in diopters) is typically less than the value expected theoretically from the inverse of the observation distance (also in diopters). This lag of accommodation is most evident at low refresh rates: the accommodative response increases with flicker frequency up to 40 Hz (Owens and Wolfe, 1985). Chauhan et al. (1992) found a further increase up to a frequency of 100 Hz. Neary (1989) reported an increased accommodation response at certain rates of intermittency, both above (50 Hz) and below (25 Hz) flicker fusion.

Kennedy and Murray (1991, 1993) and Wilkins (1986) investigated the possibility that natural saccadic eye movements (during steady illumination or at very high modulation frequencies) may be adversely affected by intermediate frequencies around 50 - 100 Hz, which are common refresh rates on CRT screens.

The present paper reports results of an experiment that investigates visual functions in subjects observing a CRT-display operated at refresh rates in the range of 50 - 300 Hz. These visual functions were accommodation, fixation disparity (the precision of convergence of the two eyes), pupil diameter, and the frequency and duration of eye blinks.

Methods

Visual functions were measured repeatedly in short test periods of 3 minutes duration while subjects looked at a CRT screen operated at different refresh rates.

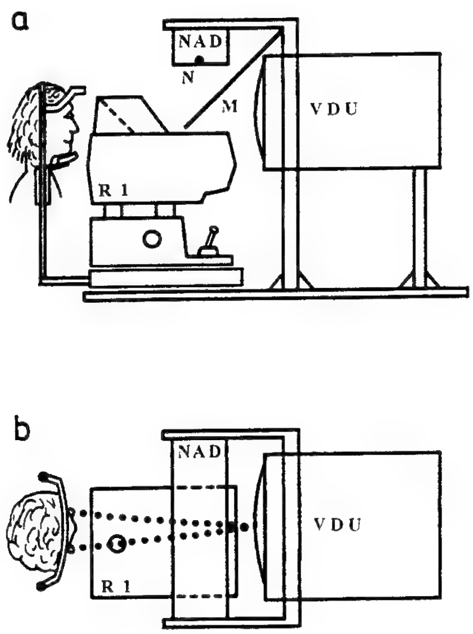


Figure 1. Apparatus with the autorefractometer (R1), the nonius alignment device (NAD), and the visual display unit (VDU). The half-silvered mirror (M) superimposes the nonius targets (N) onto the fixation charaters on the VDU.

The apparatus is shown in Figure 1. The visual target was generated on a monochrome CRT-screen at 50 cm viewing distance. An autorefractometer CANON-R1 (which automatically measures the state of refraction of the eye) was placed between the subject and the CRT screen. This system allows the measurement of accommodation while the subject has an unrestricted view of the CRT-screen. Pupil size and eye blinks were evaluated from the video image of the eye that is provided by the autorefractometer. We also measured to what extent the convergence angle of the eyes (between the two lines of sight) was properly adjusted to the stimulus, or whether small misalignments (fixation disparities) occurred so that the point of fixation is not projected onto the center of the fovea

(Jaschinski-Kruza, 1994). These small errors in convergence typically amount to a few minutes of arc and have been associated with asthenopic complaints. We measured fixation disparity by presenting two nonius test targets dichoptically, i.e., one to each eye, by means of polarizing filters. The horizontal position of the targets was variable. In case of fixation disparity the targets must have a physical offset in order to be imaged onto the center of the fovea in each eye and to be perceived as being aligned. The nonius targets were produced by a nonius alignment device (NAD) and optically superimposed onto the VDU image with a half silvered mirror. The VDU was purpose-designed in order to be able to present different frame frequencies in the range of 50 - 300 Hz. The bright screen area of 50 cd/m2 was 20 deg wide and 14 deg high, and contained an area of black numbers. Due to the fast phosphor of the CRT, full temporal modulation was maintained at frequencies up to 100 Hz, while at 300 Hz modulation was about 65%.

Experiment 1

Subjects viewed with both eyes. In order to cover the range of typical VDU repetition rates, we compared 300 Hz with 50 Hz; the latter gave visible flicker for most subjects. The results showed no effect of repetition rate on accommodation (with binocular vision) or on fixation disparity. However, pupil diameter was 0.055 mm smaller at 50 Hz ($p < 0.05$, Wilcoxon test). In order to test the reproducibility of the effect within the session, we made a separate data analysis for the two halves of the session. In the first and second half of the session the difference in pupil size between the refresh rates was 0.046 + 0.117 mm ($p = 0.099$) and 0.049 + 0.092 mm ($p < 0.044$), respectively. (The average of these two differences does not correspond to the value of 0.055 mm reported above for the complete session, since different baselines had to be used for these two analyses.) Although these differences on the basis of half of the session were only moderately statistically significant, they were highly correlated ($r = 0.89$; $p < 0.001$). In some individuals, the difference was considerably greater than the mean. Interestingly, the subject with the strongest effect (0.31 mm) showed a similar effect of 0.35 mm in a repeated session but did not report perceiving flicker in the 50 Hz condition although asked repeatedly.

Experiment 2

Subjects viewed the screen with one eye so that convergence-induced accommodation could be ruled out. The 300 Hz-condition was compared with the lowest repetition rate that did not produce visible flicker for each subject. These fell within the range of 55 - 90 Hz (mean 70 Hz). At the lower frequency, mean accommodation was 0.06 D weaker (n=17, $p < 0.05$), the median eye blink duration was 6% shorter (n=23, $p < 0.05$) and the mean eye blink interval was 15% longer (n=23, $p < 0.05$). In this test, change of pupil size was insignificant across the group.

A part of the group was retested in Experiment 2. We chose those subjects who had earlier shown an individually-significant effect in pupil size or accommodation. For the effect of non-visible flicker on accommodation in monocular vision we found a significant test-retest correlation of $r = 0.75$ ($p < 0.01$; $n = 10$) between test 1 and test 2.

The results in the two experiments were different in that accommodation was affected in monocular vision, but not in binocular vision. This difference could have been produced by the component of accommodation that is induced by convergence: in binocular vision the activity of convergence could have supported accommodation via the coupling of these two oculomotor mechanisms. This would mean that under the natural, binocular, viewing condition, accommodation was unaffected by intermittency. However, the natural coupling between accommodation and convergence may be disturbed.

Conclusion

The present study shows reproducible effects of refresh rate near critical flicker frequency on pupil size and monocular accommodation. The amount of these effects differed among the subjects. These results and the studies reviewed above demonstrate that several visual functions can be affected by frequencies of intermittency that are typical of VDU workplaces, in some cases under conditions where flicker was not visible. However, not all research in this field provides a physiological explanation of the observed effects since these tend to be influenced by the specific visual task involved, the actual viewing conditions, and the individual subject.

Possible contributions of intermittent light to visual discomfort were first investigated when fluorescent luminaries were introducted at workplaces, and visual complaints and incidence of headache were reported by part of the employees. Earlier studies were reviewed by Brundrett (1974) who concluded that intermittent lighting could be fatiguing, but that the magnitude was small and could be masked by general fatigue. From Wilkins et al. (1989), Padmos (1988) and Lindner (1994) it can be concluded that there is some evidence that 100 Hz intermittency of fluorescent light raises the incidence of headache and visual fatigue, and that individual differences may play a role. According to Lindner and Kropf (1993), those subjects who complain more than others about fluorescent lighting tend to have the following individual characteristics: they are predominantly female, aged 20 - 30 years, and have a higher psychovegetative lability, diminished power of concentration, enhanced light and flicker sensitivity, and reduced binocular and stereoscopic vision. Subjects who attribute their complaints to intermittent light might be helped by a refresh rate as high as possible, by lower luminances, by a dark background, or by LCD screens.

References

Berman, S.M., Greenhouse, D.S., Bailey, I.L., Clear, R.D. & Raasch, T.W. (1991). Human electroretinogram responses to video displays, fluorescent lighting, and other high frequency sources. *Optometry and Vision Science, 68*, 645-662.

Brundrett, G.W. (1974). Human sensitivity to flicker. *Lighting Research and Technology, 6*, 127-143.

Chauhan, K., Charman, W.N., Halnan, A.M., Kelley, C.M., Loughlin, A., Neilson, K.J. & Walsh, G. (1992). Time-averaged accommodation response to flickering stimuli. *Ophthalmic and Physiological Optics, 12*, 327-334.

Jaschinski-Kruza, W. (1994). Precision of convergence of the eyes while viewing a visual display unit. In H. Luczak, A. Çakir & G. Çakir (Eds.), *Work with display units 92*. Amsterdam: Elsevier.

Kennedy, A. & Murray, W.S. (1991). The effects of flicker on eye movement control. *The Quarterly Journal of Experimental Psychology, 43A*, 79-99.

Kennedy, A. & Murray, W.S. (1993). 'Flicker' on VDU screens. *Nature, 365*, 213.

Lindner, H. (1994). Asthenopische Beschwerden bei Beleuchtung mit Gasentladungslampen (GEL). *Ophthalmologe, 91*, 201-211.

Lindner, H. & Kropf, S. (1993). Asthenopic complaints associated with fluorescent lamp illumination (FLI): The role of individual disposition. *Lighting Research Technology, 25*, 59-69.

Neary, C. (1989). The effect of high frequency flicker on accommodation. *Ophthalmic and Physiological Optics, 9*, 440-446.

Owens, D.A. & Wolfe, J.M. (1985). Accommodation for flickering stimuli. *Ophthalmic and Physiological Optics, 5*, 291-296.

Padmos, P. (1988). *Gezondheidsschade door TL-Verlichting?* Een Literatuurstudie. TNO-Report (C-31). TNO Institute of Perception, Soesterberg (NL).

Sternheim, C.E. & Cavonius, C.R. (1972). Sensitivity of the human ERG and VECP to sinusoidally modulated light. *Vision Research, 12*, 1685-1695.

Wilkins, A. (1986). Intermittent illumination from visual display units and fluorescent lighting affects movements of the eyes across text. *Human Factors, 28*, 75-81.

Wilkins, A.J., Nimmo-Smith, I., Slater, A.I. & Bedcos, L. (1989). Fluorescent lighting, headaches and eyestrain. *Lighting Research and Technology, 21*, 11-18.

Research on simulator-based training and instruction strategies

John C.G.M. van Rooij
Training and Instruction Group
Department of Skilled Behaviour
TNO Human Factors Research Institute
The Netherlands

Abstract

Modern technology offers many possibilities for measuring, logging, and processing trainee performance data and for using this information in optimizing and automating the progression of training scenarios and the delivery of instruction. However, most decisions with respect to training and instruction are made subjectively and on an intuitive basis. The conceptual framework described in this paper guides research on the development of objective methodologies for analyzing and optimizing strategies for simulator-based training and instruction. The general approach is illustrated by results from two recent studies: one dealing with the optimization of part-task training strategies and the other dealing with the automated delivery of instruction.

Introduction

The Training and Instruction Group at TNO-HFRI performs research and consultancy in training and instruction, particularly in those areas in which advanced training media are used. The research activities of our group can be divided into strategic research and applied research projects. Applied research projects comprise both military and industrial projects. Our strategic work consists of long-term projects aimed at the development of expertise, methodologies, and tools, and is organized in four different types of tasks or domains, viz. team tasks, cognitive tasks, procedural tasks, and high performance tasks. The focus of this paper is on high-performance tasks.

High-performance tasks are complex, time-critical, steering and control tasks in which the operator is in the primary control loop of the system (cf. Schneider, 1985). An example is piloting a combat helicopter. The time-critical aspect derives from the fact that the to-be-controlled system is dynamic and operates in a dynamic and often hostile or dangerous environment. The complexity of these tasks arises from the number of, the variety of, and the interactions between task components which, apart from perceptual-motor components, typically also comprise (subsidiary) procedural and cognitive components.

One of the training characteristics of these tasks is selection which is often required because many people fail to develop proficiency. Even after selection, the

duration of training required to reach an operational level of performance may be considerable. Typically there are large differences between novice, advanced, and expert operators, not only with respect to the speed and accuracy of performance but also with respect to the use of different strategies. Training usually involves a part-task or training scenario since training in the operational environment is often dangerous, expensive, or impossible.

Not surprisingly, with advances in simulation technology, an increasing amount of training and instruction is provided in simulated training environments. Training and instruction in a simulated training environment offers several advantages over training in the operational environment: e.g., lower cost, less risk, and better and more varied opportunities for learning. These opportunities offer the possibility to increase the number of learning experiences per unit of time, the possibility of arranging training conditions to fit particular training needs, the opportunity for detailed and objective performance measurement, and opportunities for standardizing and automating training and instruction strategies.

However, despite these opportunities, current practices for training and instruction are still usually modelled after the way training and instruction is delivered on the operational environment, viz. apprenticeship instruction (Schank and Jona, 1991). Although such an approach has a high face validity, it wastes resources, and depends on the teaching abilities of the instructors. At any rate this traditional approach does not exploit the opportunities offered by technology to substantially increase training effectiveness.

Most studies in training and simulation reflect a one-sided concern with issues of fidelity and transfer of training instead of issues associated with the effectiveness of alternative training and instruction strategies and the possibilities for rendering these strategies more efficient, e.g. by automating them. The latter issues are the concerns of one of our strategic projects. The general framework of this project, viz. the Training and Instruction Model (TIM), is described in the next section. In subsequent sections, the results of two studies that were conducted within the TIM-framework are described briefly. The final section of this paper concludes with a summary of the main findings and points to potential applications.

The training and instruction model (TIM)

TIM is intended as a general framework for conducting research to identify better simulator-based training and instruction strategies (Van Rooij, 1994). The purpose of TIM is to render the problems of simulator-based training and instruction more tractable and amenable to analysis. The scope of TIM is constrained by its focus on skill acquisition in the domain of high-performance tasks. Of course, the domain of high-performance tasks in itself represents a unlimited variety of different tasks. However, we believe that with respect to training and instruction strategies there is much more commonality across these tasks than there is commonality in task content. Our approach therefore consists of starting from global formal commonalities with respect to training and instruction characteristics and then working our

way down to specific applications, rather than the more conventional method of developing a training and instruction model for a particular task and subsequently attempting to generalize it to other tasks.

Within the context of TIM, learning and transfer (and retention) are treated as dependent variables that can be optimized by manipulating the independent variables represented by training and instruction parameters. The problem in devising training and instruction strategies that together make up a training programme is in selecting appropriate training and instruction parameters and in assigning values to those parameters in a way that optimizes training effectiveness.

There are several criteria that can be used to assess and/or optimize training effectiveness, viz. end-of-training, transfer, and retention criteria. Criteria can be further subdivided into those that are assessed in terms of performance level at one or more moments of time (time-referenced) or criteria that are assessed in terms of amount of training time at one or more level(s) of performance (performance-referenced). The criterion that is most useful depends on the goal set for training and on the constraints that are imposed.

In conducting research within the context of TIM two complementary approaches have been followed: (1) a theoretical-experimental approach in which concepts and methodologies are developed and tested in a generic training and instructional environment and (2) an applied-empirical approach in which generic guidelines and tools derived from the theoretical-experimental work are applied in the practical development of training programmes.

The generic training and instruction system (TIS) that is developed for the theoretical-experimental work is build around the Space Fortress Game (SFG), a PC-based computer game. The SFG was specifically designed for research on training and instruction strategies (Donchin, 1989; Mané and Donchin, 1989) and both the features of the task and the associated training characteristics are considered to be representative of high performance tasks.

SFG consists of three main part-tasks: ship control, mine handling, and resource management. SFG is played in 5-minute games. The main goal in playing the SFG is to maximize the game score (= the number of points acquired during a single game). Points can be earned primarily by shooting at and destroying an enemy space fortress by controlling a spacehip by means of a joy stick. Two experimental studies that were conducted within the context of the SFG-TIS. will be described.

Experiment 1: adaptive part-task training

The issue addressed by the first experiment was: "given a fixed amount of training time, i.e. a time-referenced criterion, and individual differences in learning level and potential, what is the effect of variation in the amount of training across part-tasks on end-of-training performance?"

In this experiment the amount of training per training phase was manipulated by varying the number of trials. Apart from the game score, performance on SFG can be expressed in terms of Fort-Destruction Times (FDTs), the time it takes to destroy the fortress. Both measures, game score and FDT, are highly correlated. Therefore, FDT was used as the definition of a trial. Because the effective training time is composed of a series of FDTs, adopting this trial-definition enables one to estimate the number of trials that fit within a given amount of time.

The rationale of the experiment is illustrated in Figure 1.

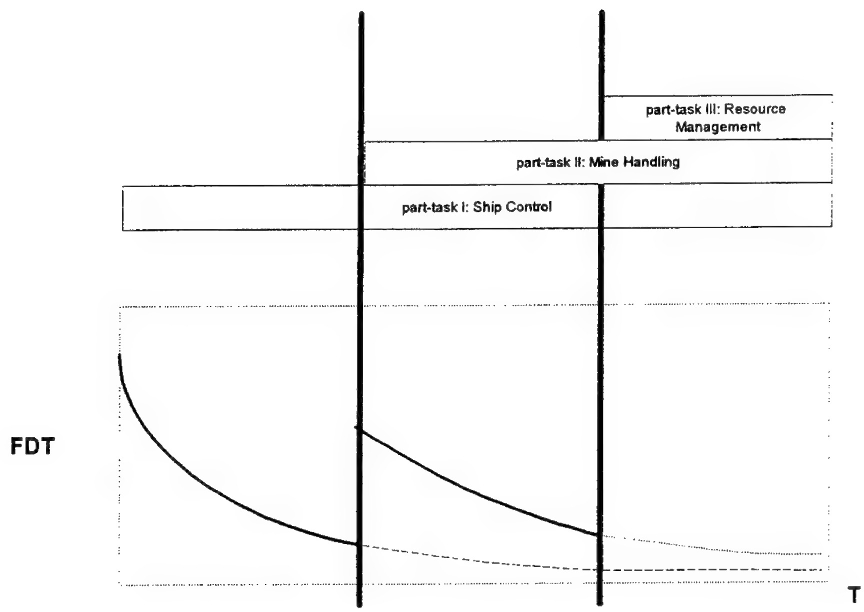


Figure 1. Rationale of experiment 1 on adaptive part-task training

Training phases were composed according to a cumulative training scheme. A cumulative training scheme is a training scheme in which successive part-tasks are incorporated in the training task one by one in a predefined order. Part-tasks were incorporated in the following order: ship control, ship control + mine handling, and ship control + mine handling + resource management. The criterion for adding a part-task, or, put differently, the criterion for being promoted to a subsequent training phase, consisted of a particular percentage of trials that had to be spent in each training phase. Percentages were expressed in terms of the estimated number of trials that were estimated to fit within the (remaining) training time. These estimates were based on the learning curves that were fitted to the training data within each training phase (for full details the reader is referred to Van Rooij and Roessingh, 1994). Two different criteria were needed: one criterion for adding mine handling and another for adding resource management. For each criterion four

different percentages were used, viz. 0, 2, 10, and 50%. Factorial combination of these two sets of percentages results in the design shown in Table 1.

Table 1: Design matrix of experiment 1 on adaptive part-task training.

↓ I-II II-III →	0%	2%	10%	50%	
0%	n = 6	n = 2	n = 2	n = 2	n = 12
2%	n = 2	n = 2	n = 2	n = 2	n = 8
10%	n = 2	n = 2	n = 2	n = 2	n = 8
50%	n = 2	n = 2	n = 2	n = 2	n = 8
	n = 12	n = 8	n = 8	n = 8	n = 36

The combination 0%/0% represents the control group (6 subjects). The subjects in this group were trained on the entire task, i.e. phase 3, from the start. The combinations 0%/2%, 0%/10%, and 0%/50% represent those groups that were trained on ship control + mine handling, i.e. phase 2, from the start and subsequently were promoted to phase 3. The combinations 2%/0%, 10%/0%, and 50%/0% represent those groups that were trained on ship control, i.e. phase 1, from the start and subsequently were promoted to phase 3 (thereby skipping phase 2). All other combinations represent different amounts of training for each phase.

All trainees received the same instruction prior to training and were trained for a total of 16 hours. End of training performance was defined as the average game score computed over the last 20 games. A multiple-regression model, i.e. a response surface, was fitted to the data where end of training performance was predicted by the number of trials per training phase.

Overall, performance of the control group was not significantly different from the performance of the other experimental groups. This implies that, on average, part-task training does not necessarily yield better results than training on the whole task. Some experimental groups performed worse than the control group and some experimental groups performed significantly better. Thus, whether part-task training results in better training results than whole-task training depends on how the available training time is allocated across part-tasks. Moreover, it can be concluded that the design of this experiment offers a method for optimizing this allocation.

Experiment 2: automated delivery of instructional interventions

As noted in the introduction, most decisions regarding simulator-based training and instruction are resolved on an ad-hoc and intuitive basis by the instructor in charge. As an alternative to this approach, a lot of effort is devoted to the development of Intelligent Tutoring Systems (ITSs) and related computerized techniques. ITSs are systems that are based on a thorough in-depth analysis of the task/domain and the

training process that is to be taught. The results of such an analysis combined with the use of techniques from Artificial Intelligence are subsequently incorporated into an ITS. So far, most ITSs described in the literature have not passed the prototype stage and most attempts have focussed on relatively well-structured cognitive tasks and domains. In contrast, much of the difficulty in modelling high-performance task are due to their dynamic, complex, and real-time nature. In most cases, the analytical approach required to develop an ITS for such tasks would be far too laborious to be feasible.

A second experiment conducted within the TIM framework was set up to investigate the potential of a statistical approach to modelling the instruction process as opposed to existing intuitive and analytical approaches. The interest in such an approach is not only motivated by the possibility it may offer for automating instruction strategies but also by the possibility that it may offer a means to objectify and study the effects of such strategies.

Due to the real-time nature of high-performance tasks, an important issue is when and how to diagnose / sample the training process. One method is to interrupt the training process at fixed intervals and to deliver instruction between intervals (interval driven). Another option is to link diagnosis and interventions to particular events, e.g. errors, that may occur during training (event driven). Finally, diagnosis and interventions may be coupled to the values of particular training process parameters, e.g. cumulative records of particular events (parameter driven).

Our experiment focussed on the first option. The rationale of the experiment is shown in Figure 2. The description will be limited to the first part of the experiment that focussed on the effect of instructional interventions of learning the ship control part-task of the SFG.

Apart from game score as an overall performance measure, for ship control, performance during each game period is described by 20 other performance measures.

During the first phase of the experiment, 6 trainees (the control group) trained for 48 games. These games were recorded which enabled a full replay of each game. Based on game replays and data plots of performance measures versus games, an expert player was asked to assign instructional interventions to each of the 288 games that had been recorded. In this task, interventions had to consist of text or text accompagnied by recorded game samples to demonstrate game tactics. Finally, 6 interventions were designed and assigned. The performance measures of the games that were recorded, together with a number that indicated the associated instructional intervention that had been assigned to it, was input to a Multiple Discriminant Analysis (MDA). A MDA yields as output a set of classification functions that, given a set of performance measures, enables the selection of that intervention that, in a statistical sense, best matches the performance measures. These classification functions constitute a statistical model of the (instruction) strategy used by the expert player in assigning interventions to games. We

hypothesized that this model can be used to assign the same interventions to newly-obtained games and, hence, as a means to automate instruction.

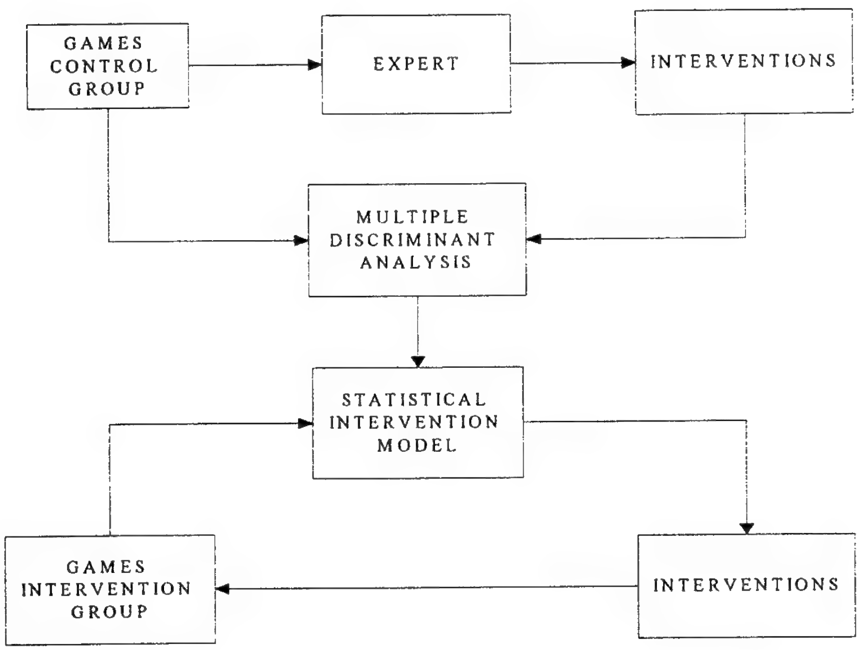


Figure 2. Rationale of experiment 2 on automated delivery of instructional interventions

During the second phase of the experiment, the power of the statistical model to provide automated instruction was tested. A second group of 6 subjects, the intervention group, was trained in the same way as the first group except that between games they received interventions that were delivered according to the statistical model derived from the MDA.

The effect of each of the six interventions was assessed by computing the difference between performance measures of the games preceding and following the intervention. These differences were computed across all trainees in the intervention group. All difference scores displayed the effects that were intended by the respective interventions.

The response of individual trainees to interventions was assessed by comparing difference scores across trainees. These comparisons revealed that trainees 5 and 6 were less responsive, i.e. they had zero or low difference scores. In particular, the performance of trainee 5 was far below expectation: his game scores were lower than those of his matched counterpart in the control group. On the basis of their scores on a selection test, both subjects had been classified as high-ability subjects.

This suggests that individual differences in ability may interact with the effectiveness of instruction.

Training curves were obtained by plotting average game scores versus game number for each group. Although, overall, the training curve of the intervention group was higher than the curve of the control group, the difference was not statistically significant. However, when the data of subjects 5 and 6, i.e. the high ability trainees who had been found to be less responsive to interventions, were excluded from both groups, the training curves displayed in Figure 3 were obtained. Although both groups were briefed in the same way on the optimal control strategy to use, figure 3 shows that the intervention group outperformed the control group on all games. Also, the variability in performance of the intervention group is lower than that of the control group. All trainees in the intervention group consistently tried to adhere to the same optimal control strategy whereas the subjects in the control group were more inclined to waver. Both curves follow the usual learning power curve. The learning rate in the intervention group, indicated by the steepness of the curve, is higher than the learning rate in the control group. Due to the curvilinear shape of training curves, small differences in performance (the y-axis) correspond to increasingly larger differences in training time (the x-axis). This means that relatively small differences in the training criterion may have large consequences for the training time required. Thus, although the effect of instruction in terms of performance may appear to be small, the effect in terms of savings in training time may be substantial.

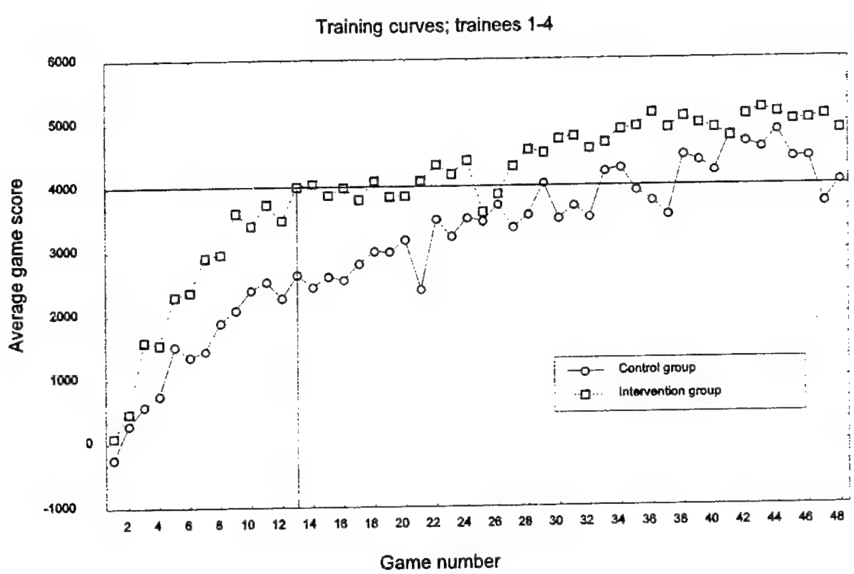


Figure 3. Average training curves for the low ability trainees (trainees 1-4) of the control group and the intervention group

The horizontal line in Figure 3 indicates a score level of 4000 game points. By drawing vertical perpendicular lines between the point where this line intersects the training curves, the corresponding savings in training time can be computed.

The intervention group (the solid vertical line) requires 13 games (65 minutes) to reach this score level. The control group (the dotted vertical line) requires 29 games (145 minutes) to reach the same criterion. Thus, for this 4000 game-point criterion, the intervention group only needs 45% of the training time the control group needs. In other words, the delivery of automated interventions results in a saving of 55% in training time (16 games).

In summary, the results of this experiment demonstrate that it is possible to automate instruction by statistically modelling the behavioral correspondences between trainees and instructor and that this may improve training effectiveness both qualitatively (use of game strategy) as well as quantitatively (higher game scores or shorter training times).

Conclusion

The TIM framework is intended to provide guidelines for designing training programme and for acquiring and analyzing training data in order to be able to subject these data and, hence, the corresponding training programme to objective statistical analysis. The results of the experiments so far show that this approach bears considerable promise as a means to investigate and to improve the training effectiveness of simulator-based training programmes.

References

Donchin, E. (1989). The learning strategies project. *Acta Psychologica, 71*, 1 - 15.

Mané, A. and Donchin, E. (1989). The space fortress game. *Acta Psychologica, 71*, 17 - 22.

Schank, R.C. and Jona, M.Y. (1991). Empowering the student: new perspectives on the design of teaching systems. *The Journal of the Learning Sciences, 1*, 7 - 35.

Schneider, W. (1985). Training high-performance skills: fallacies and guidelines. *Human Factors, 27*, 285 - 300.

Van Rooij, J.C.G.M. (1994). *Training and instruction strategies for complex time-critical steering and control tasks in simulated training environments.* Report TNO-HFRI (in preparation), Soesterberg, The Netherlands.

Van Rooij, J.C.G.M. and Korving, F. (1994). *Automation of simulator-based instruction.* Report TNO-HFRI (in preparation), Soesterberg, The Netherlands.

Van Rooij, J.C.G.M. and Roessingh, J.J. (1994). *Training transfer and the optimization of part-task training strategies.* Report TNO-HFRI (in preparation), Soesterberg, The Netherlands.

Participating in simulations –good, middling and irrelevant

Helen E.Ross[1] *& A.A.I.Wedderburn*[2]

[1]*Department of Psychology, University of Stirling, Stirling FK9 4LA, Scotland*
[2]*Department of Business Organisation, Heriot-Watt University, Edinburgh EH14 4AS, Scotland*

Abstract

This paper describes simulations in which the first author participated as a learner or observer, and evaluates them from a participant's viewpoint. The best training situations were simulations of Spacelab Missions run by NASA and ESA for the benefit of the experimenters, astronauts and managers involved. Parabolic flights were also useful both as simulations of zero gravity in orbital flight, and in providing experimental data. A poorer training was the simulated emergency evacuation of a passenger ferry, since much behaviour was unlike a real emergency; but some useful lessons were learned. A simulated evacuation of an aircraft was also useful for experimental purposes but had little training value. A corporate training weekend for a large organisation involved participation in irrelevant simulated tasks and games: it seemed bizarre and pointless.

Introduction

Simulations and training sessions are run for a variety of reasons, and their success may be evaluated differently by managers and participants. Landy (1989) describes the purpose of simulations as "to gain the control that may be absent in a field experiment but at the same time to approximate a realistic operating situation so that one can generalise from the research findings to the operational task. The key word here is realistic .." (p.30) In terms of training, simulations aim to bridge the gap between efficient knowledge acquisition and transfer to the situation on the job. "A simulation seems to be an ideal compromise that combines the best of both techniques" (Dipboye et al. 1994). In practice, many implementations are what Goldstein (1991) calls "part-simulations, which replicate a critical or difficult portion of the task without attempting to provide a complete environment".

The best events are carefully prepared and their purpose is transparent to all parties. The following events are ones in which the first author (HER) participated as a learner or observer, so her evaluations probably differ from those of the management or experimenters.

Good simulations - NASA and ESA

Spacelab Simulations

Good simulations were run by the space agencies NASA and ESA, which have a lot of experience in training their participants. HER was involved in running experiments on the Spacelab 1 mission in 1983 (a joint NASA and ESA venture) and the Spacelab Dl mission in 1985 (a joint NASA and German DLR venture). Much time was usefully spent in training the astronauts to run individual experiments, but in addition all the experimenters were required to participate in a simulation that lasted for a few days. The simulation was designed to follow part of the 'time-line' of the real mission, and the experimenters were required to go through the activities that would probably occur in practice. They were trained in the use of the voice communication system and the acceptable style of messages, keeping a log of events, sending faxes and other messages, dealing with emergencies, and working with the same personnel as in the real mission. Much of this training was essential, and most was useful in enabling the real mission to run smoothly. Participants could see the value of the exercise, and all parties thought it was worthwhile.

Parabolic Flights

Another useful activity run by NASA and ESA is parabolic flights. These flights follow repeated parabolas in each of which there is about 20 seconds of near zero-gravity, preceded and followed by accelerations of up to 2 *g* (Pletser, 1989). The 0 *g* phase approximates the weightless condition of spaceflight, and can be used to train astronauts before a space mission. It can also be used to try out and perhaps modify some potential spaceflight experiments (Frimout & Gonfalone, 1985). HER's experiment was originally planned from inadequate ground-based simulations, but she was able to improve it as a result of experiments in parabolic flight (Ross, 1981, 1985). Both astronaut training and preliminary experiments are useful aims. In addition, parabolic flight experiments are often very successful, and are publishable in their own right regardless of any future space experiments (e.g. Ross & Reschke, 1982). Thus parabolic flights occupy a position that is sometimes regarded as a simulation or training, and sometimes as the real thing (Pletser, 1989).

It should be noted that participation in parabolic flights is such an exacting activity and that medical examinations and physiological and safety training are required before one is allowed to take part (e.g. Lapinta 1982). HER has participated in various such training sessions in the USA, England and Germany, and found them all useful and well-organised .

Middling simulations - evacuations

Evacuations are sometimes used to train participants (as in fire drills), and sometimes as quasi-experiments to discover what might go wrong or as real

experiments to measure some variable. They usually suffer from a lack of realism, though they may have some merit in other respects.

Evacuation of a passenger ferry

Exercise Claymore (3.10.93) was planned to test the co-ordination of the rescue services in a ferry emergency. The simulated emergency was a fire on a CalMac passenger and car ferry (MV Claymore) in the Firth of Clyde (H.M. Coastguard, 1993). HER attended as an observer, to look at another area of concern –the behaviour of the passengers and their interaction with the ship's crew. The simulation was not ideal for that purpose, for the reasons given below.

Composition of Passengers
The 'passengers' for this exercise were over 150 members of the Territorial Army (52nd Lowland Volunteers, TA) and about 30 members of the British Red Cross. The TA consisted mainly of healthy young males, though there were a few young women. The Red Cross covered a wider age range, and probably contained more women than men. Several members of the Red Cross were dressed to simulate injured persons. There were also 15 heavy dummies on board, representing unconscious or dead persons. This sample of people cannot, of course, represent a typical mixture of ferry passengers. A normal mix would probably contain a much wider age range, with many family groups and older people, and possibly parties of school children or other youth groups. Infirm elderly people and young children would require assistance in entering life boats. Passengers would be anxious and some might become hysterical. Family groups would attempt to stay together.

General Behaviour of Passengers and Crew
Naturally, none of the above behaviour occurred during the exercise. The volunteer passengers were calm and relaxed, enjoying a Sunday outing in good weather. People waited around chatting to each other amicably, to see what would happen next. The crew and all other participating groups took a similarly relaxed attitude. The evacuation seemed to proceed very slowly, and would probably have been much quicker in a genuine emergency.

Information and Crew Operations
There was one respect in which the exercise simulated genuine emergencies: lack of accurate information to the 'passengers' about what was happening or would happen (Kuo et al, 1992; Kennedy, 1993). While the observers were very well briefed about the organisation of the day, many volunteers were given conflicting information or very little information. The Red Cross 'badly injured' were at first told that they were to be helicoptered off the vessel, as it was easier to do that than to get them on the lifeboats; they were then told that they were not insured for the helicopter, and would be moved to the boats after the sound or 'walking wounded' passengers. TA volunteers were helicoptered off instead of Red Cross volunteers. The volunteers in the assembly area (D deck) said that no public address (PA) system was used for announcements, or if it was it could not be heard. Instructions

for entering the lifeboats were unclear, and changed from the injured first to the uninjured first.

Instructions were given by CalMac personnel, who were too hesitant and polite. As one volunteer put it: "If they can't handle 'good' passengers, how would they manage in a real emergency?" They thought the crew should be clearer and more authoritative. The announcers also used nautical jargon such as 'Starboard', which would need translating for the benefit of average passengers. Announcements were made only once and were often not heard: they should be repeated firmly several times. The CalMac representative demonstrating the use of lifejackets was apparently unable to put his own on at the first try. There were no instructions as to what to do in the water, such as the use of whistles. Passengers were told to read the lifejacket instructions on the wall. The paramedics asked the passengers to help move the injured, but did not give any instructions about taking care. The issuing and counting of boarding passes was haphazard, and it is not surprising that the count of passengers did not tally.

Conclusions
Exercise Claymore proved a useful occasion for observing problems in passenger-crew interaction, even though that was not it's main purpose. It showed that the crew required further training in coordinating their activities, handling passengers, and giving out clear and useful information.

Evacuation of an aircraft

The Applied Psychology Unit at Cranfield has run a series of cabin evacuations, to investigate speed of evacuation under various different conditions such as the configuration of the seats and exits, the presence of smoke, or (in this case) the use of internal water sprays. These experiments have provided much valuable data (Muir et al., 1989, 1990; Muir & Bottomley, 1992). However, the problem of lack of reality remains. The volunteers were aged between 20-50 years, and were healthy. They knew that an evacuation would occur, and that water might be used. There was no sense of panic, only a slight disinclination to get wet. However, a recording of screams was played, and at the time HER thought this was real. A video of the evacuation was played back afterwards, and it was interesting to compare the video with the evacuation experience.

Evacuations of this sort are certainly useful for their intended experimental purpose, but probably provide little in the way of training for the participants.

Irrelevant simulations - corporate training

Corporate training or management training exercises are fashionable at present. Personnel are sent sailing or hillwalking together, or asked to take part in group games, in the belief that they will work better as a team when back at the office.

HER was a member of a regional board of the Nature Conservancy Council for Scotland (NCCS), which was about to be merged with the Countryside Commission for Scotland (CCS) to become Scottish Natural Heritage (SNH). Any merging of organisations is likely to be difficult, particularly when the two have some conflicting aims: in this case NCCS was supposed to conserve 'nature' against various human and other encroachments, while CCS was more concerned with human access to the countryside. Merging the two might create something like the 'pushmi-pullyu' - a mythical animal with two heads pulling in opposite directions (Lofting, 1922).

To ease the transition, people connected with both organisations were sent to a hotel in the Highlands for three days of corporate training. Various welcoming speeches were made, but only a minimal explanation of the purpose of the gathering was given - ending "It's about you - getting to know yourself and others." The next day participants were put into groups of about 12 and made to play games by the training leaders from a Scottish college. The leaders were mainly young women, who explained the rules of each game but not the ulterior purpose. The first task was to calculate the day of the week on which a construction job would be completed, given various items of information. Members read out their information inaudibly against a lot of background noise (I could not hear what was said). The men did most of the calculating, while the women kept silent. Next there was a construction task, in which the aim was to build a stand as high as possible out of newspaper and sellotape: my group excelled at this. The group was then required to discuss the order in which castaways should be rescued from a cave, knowing that the later ones would probably die, and given only scanty biographical material about the people. The game had obviously been imported from England: only one castaway had been born in Scotland, and all the rest in England. My team refused to play the game seriously, and decided to rescue the Scot first and draw lots for the rest. The trainer was rather bemused by this behaviour, and said the group was much more decisive than most students. The next games were held outdoors: group members climbed on each others' backs to put a tyre over a high post; lay on the ground and stretched themselves out to reach a target while forming a linked chain; and walked in groups of four on simulated 'ski' planks. Finally the group had to use its earnings from the previous games to 'buy' planks and other objects to cross over on to an island. My group was not much good at this. I found these unreal games rather boring. I also felt frustrated during some supposedly real activities in the evening - curling and country dancing - since these were run in a shambolic manner and were not taken seriously by the other participants. I left early the next day.

It is not clear what was achieved by these 'training' games that could not have been achieved more cheaply by real games or a night in the pub. Corporate training exercises serve a financial purpose for those who sell the courses; they perhaps give satisfaction to managers who feel they have done their best to improve the morale of their workforce; and some of the workforce may enjoy a paid holiday. But there seems to be little scientific evaluation of the efficacy of such exercises. In the case

of SNH, the training has not solved the "pushmi-pullyu" problem, since the organisation continues to have a reputation for giving out contradictory messages (McOwan 1994).

Conclusions

Simulations can provide useful training if they are well prepared and are relevant to the purposes of individuals or groups. Low-fidelity simulations, with lack of task and response realism, can still provide moderate predictive validities that make them cost-effective (Motowidlo & Tippins, 1993). Buying irrelevant training packages from outside vendors is unlikely to be useful.

References

Dipboye R.L., Smith C.S. & Howell W.C. (1994). *Understanding industrial and organizational psychology.* Fort Worth: Harcourt Brace College Publishers.

Frimout, D. & Gonfalone, A. (1985). Parabolic aircraft flights - An effective tool in preparing microgravity experiments. *ESA Bulletin 42*, 58 - 63.

Goldstein, I.L. (1991). Training in work organizations. Chapter 9 in M.D. Dunnette and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology.* Palo Alto: Consulting Psychologists Press.

H.M. Coastguard (1993). *Exercise Claymore.* 3rd October 1993. Glasgow: Strathclyde Emergencies Co-ordination Committee.

Kennedy, B. (1993). *A human factors analysis of evacuation. escape and rescue from offshore installations.* Offshore Safety Division of the Health and Safety Executive. Offshore Technology Report - OTO 93004.

Kuo, C., Ross, H. & Tennant, S.M. (1992). *Passenger behaviour analysis in emergency situations.* Report for the Marine Directorate of the Department of Transport. Glasgow: Marine Technology Centre, University of Strathclyde.

Landy, F.J. (1989). *Psychology of work behavior.* Pacific Grove: Brooks/Cole

Lapinta, C.K. (1982). *Aerospace physiological training outline.* NASA Johnson Space Center, USA: JSC-03890.

Lofting, H. (1922). *The story of Doctor Dolittle.* London: Jonathan Cape.

McOwan, R. (1994). Pace of change. *Scotland on Sunday* Nov.6th, p.21.

Motowidlo, S.J. & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organisational Psychology, 66,* 337 - 344.

Muir, H. & Bottomley, D. (1992). *Aircraft evacuations: a preliminary series of aircraft evacuations to investigate the influence of acoustic attraction signals located beside the exits.* Civil Aircraft Authority Paper 92002.

Muir, H., Marrison, C. & Evans, A. (1989). *Aircraft evacuations: the effect of Passenger motivation and cabin configuration adjacent to the exit.* Civil Aircraft Authority Paper 89019.

Muir, H., Marrison, C. & Evans, A. (1990). *Aircraft evacuations: Preliminary investigation of the effect of nontoxic smoke and cabin configuration adjacent to the exit* Civil Aircraft Authority Paper 90013.

Pletser, V. (1989). *Microgravity during Parabolic flights with Caravelle. ESA Users' Guide*. ESTEC, Noordwijk, Netherlands: ESTEC/GP/O9/VP, Issue 1, April 1989.

Ross, H.E. (1981). Mass discrimination during weightlessness: Plans for a Spacelab experiment. *21-Convegno Internazionale Scientifico Sullo Spazio*, RIENA, Rome, pp. 173 - 182.

Ross, H.E. (1985). Mass-discrimination; The development of a low-technology self-test procedure for space experiments. *Earth-oriented applications of Space Technology, 5*, 95 - 99.

Ross, H.E. & Reschke, M.F. (1982). Mass estimation and discrimination during brief periods of zero gravity. *Perception and Psychophysics, 31*, 429 - 436.

Performance measurement in driving simulators

J.E. Korteling and K. van den Bosch
TNO Human Factors Research Institute
P.O. Box 23, 3769 ZG Soesterberg
The Netherlands

Abstract

This paper is the result of a study aimed at improvement of Performance Measurement and Feedback (PMF) systems in driver-training simulators, and thereby formulation of guidelines for the development of these systems. First the major shortcomings of some existing PMF systems will be reported. These are characterized mainly by a lack of application of knowledge concerning the driving task and the way student drivers learn perceptuomotor skills. More important for the present purposes, however, is the manner in which relevant knowledge may be implemented in a PMF system for a driving simulator. Therefore, five principles that are crucial for a successful development of PMF systems for training simulators will be presented. These principles refer to the validity of the simulator for different subtasks, the relevance of subtasks for the training, the relevance of measured variables for subtasks, the manner of metric construction, and the comprehensibility of scores. In the design of a PMF system these principles should be applied systematically and in a stepwise manner. This was accomplished for two driving simulators of the Dutch Army. The global characteristics of these systems will be briefly presented and discussed.

Introduction

For the training of tracked-vehicle drivers (Leopard 2 and YPR-765) of the Netherlands Royal Army, two full-scale driving simulators were developed. These simulators include, among other things, a computer-generated and collimated image, a six degrees-of-freedom moving-base system and an instruction panel.

In order to enhance the instructor's efficiency, both simulators also are equipped with a so-called 'Performance and Marking' system, developed by the manufacturer. This is a Performance Measurement and Feedback (PMF) system that measures driving performance. Training with such PMF systems may provide two major advantages above usual training on a driving simulator: explicit feedback to the student and more objective performance judgements by the instructors (Korteling, 1990a). Feedback is of primarly relevance for the student, who needs knowledge of results (Adams, 1979, 1987; Schmidt, 1975, 1988), and objectivity is primary relevance for the instructor, who wants to compile an objective appraisal of the strong and weak points of a student's driving behaviour. These advantages are closely related. Objective performance data, for example, enable the instructors to

improve the quality of their instruction, which in turn implies that knowledge of results (for the student) is enhanced.

During our first experience with the Performance and Marking system we noticed that the large quantity of detailed output was not easy to comprehend and lacked significance for driver training. The TNO Human Factors Research Institute was therefore asked by the Royal Netherlands Army to evaluate the system and to make recommendations for improvement. The original version of the Performance and Marking system will be described and shortcomings of the system will be outlined. Furthermore, a design of a more appropriate and user-friendly system for performance measurement and feedback will be presented. This PMF system is based on general theoretical principles combined with existing knowledge of the driving task (Korteling, 1990b; Korteling & Padmos, 1990). Both the critique of the original, and the design of a new system, will proceed according to five principles that are crucial for a successful development of automated performance evaluation and feedback systems for training simulators.

Because PMF systems for training simulators have been developed only recently and therefore only limited knowledge concerning maximization of their efficiency still is available, this paper and the more extended reports (Korteling, 1990a; Korteling, 1991; Korteling & Padmos, 1992) should be considered as a first step for improvement of the effectiveness of these kinds of systems.

The Performance and Marking system

This section gives a brief description of the main characteristics of the original PMF system: the Performance and Marking system as developed by the simulator manufacturer.

Feedback of the Performance and Marking system consists of a pattern of scores on predefined aspects of driving behaviour related to objective criteria. In its original form the system monitors route driving, consisting of road and terrain driving, and what may be called obstacle driving, i.e., water wading, driving onto a low loader, over ditches, over solid blocks a "step up" or a "sloping block", etc. The last two obstacles refer to a concrete object with vertical sides or steep sloping sides, respectively, including a traverse.

The sets of performance measures, monitored by the Performance and Marking system, for route driving and obstacle driving are different. For route driving, mean and/or peak values or frequencies are measured (Fig. 1). The route driven is divided into normal (straight or curved) sections and junctions. For each single section of normal road or junction, all these variables are separately measured, stored and presented. Since a route, which usually consists of many of these sections, is intended to take about 5 minutes to drive, the corresponding output will often be huge, with printouts exceeding a meter for only one Performance and Marking evaluation.

DRIVEN OBSTACLE/ROUTE — Database: Vlasakkers — Total mark: 63%

PROGRAMMED OBSTACLE/ROUTE — Student time: 4.39 — Instructor time: 5.00

Route straight	Mark		Step: 61%		Time: 0.08				Step: 100%		Time: 0.04	
	function		mean	maximum	number	mark	function		mean	maximum	number	mark
	speed	kph	14.0	26.9	x	0	speed	kph	26.1	33.1	x	12
	RPM	*100	17.72	23.37	x	6	RPM	*100	18.62	21.15	x	12
No crash 2	acceleration	%	42	75	x	6	acceleration	%	40	100	x	12
	wheel	%	x	100	2	12	wheel	%	x	100	2	12
	brake	%	14	53	x	0	brake	%	7	22	x	12
	slip	x	x	x	0	6	slip	x	x	x	0	6
	slide	x	x	x	0	6	slide	x	x	x	0	6
	deviation	m	x	x	x		deviation	x	x	x	x	x
	indicators	#	x	x	0	8	indicators	x	x	x	0	8
	verge	#	x	x	0	8	verge	x	x	x	0	8
	suspension	%		40	x	3	suspension	%		38	x	6
	gear	#	VA	V1	x	4	gear	#	VA	VA	x	4
							crash	x	x	x	x	2

Sloping block	Obstacle mark: 55%	Time taken: 0.33	Obstacle mark: 100%	Time taken: 0.41
Approach	good speed 3.0 no brake good gear V2 good accel very fast 2.4 hard bang to suspension	17 good heading 0 60 good steer 0	speed: 3.0 kph brake: 100% gear: V2 accel: 59% speed at slope: 1.1 kph	heading: 0 deg steering: 0
Ascent	good speed 2.2 no brake good gear V2 no accel very slow 2.0 low pitch smooth ride	0 poor heading 358 0 no steer 0 29.59	speed: 2.3 kph brake: 100% gear: V2 accel: 74% speed at cofg: 6.5 kph	heading: 351 deg steering: 32 pitch rate: 41.72d/sec
Crossing	very fast 7.6 hard brake good gear V2 over accel very fast 7.5 good pitch hard bang to suspension	100 aver heading 358 80 good steer 0 45.04	speed: 4.3 kph brake: 100% gear: V2 accel: 76% speed at edge: 3.6 kph	heading: 2 deg steering: 0 pitch rate: 46.09d/sec
Descent	very fast 9.4 hard brake good gear V2 no accel very fast 5.2 low pitch smooth ride	100 good heading 1 0 no steer 0 38.77	speed: 5.4 kph brake: 0% gear: V2 accel: 39% speed at bottom: 4.2 kph	heading: 0 deg steering: 32 pitch rate: 76.28d/sec
Drive off	good speed 3.3 good brake good gear V2 no accel smooth ride	100 good heading 1 69 good steer 0	speed: 3.1 kph brake: 100% gear: V2 accel: 80%	heading: 1 deg steering: 0

Fig. 1. Partial prints (two sections) for Route and Obstacle driving of the *Performance and marking* system of the YPR-765 and the Leopard 2 driving simulators, as developed by the simulator manufacturer.

For obstacle driving, assessment of performance on the different measures is more qualitative, such as very fast, good gear, hard bang to suspension, or poor heading. These variables are separately measured at critical moments (e.g., first contact) of the different phases in which the obstacles are crossed. These phases are: approach, ascent, traverse, descent, and driving off.

Driving behaviour is evaluated by relating the student's scores on a given trajectory to the results of one expert driver (the expert database) over the same trajectory. Fig. 1 shows the heading and a partial print of a student's driving performance (left part) on a section of straight road (upper part, representing 33 s driving) and across a sloping block (lower part, representing 8 s driving), both related to an expert's (instructor) performance (right part). With respect to route driving the student's performance on each measure is marked by the degree of similarity to the expert performance and the maximum possible mark, ranging from 2 to 12. Measures regarded as important have a higher maximum mark (e.g., speed:

12) than measures regarded as less important (e.g., gears: 4). Metrics for similarity to expert performance are very arbitrary and lack a sound psychometric basis.

For each route-driving measure, the expert database determines the performance leading to a maximum mark. The sum of the student's marks for all measures within a section of the route (straight/curved, junction) or the obstacle (approach, ascent, traverse, descent, or drive off) is expressed as a percentage showing how close the student comes to the criteria in the expert database. Two points are added when there are no crashes detected during a section of the Performance and marking route. The marks sum to a compound mark of 100% when a student's driving is the same (within the minimum ranges) as the expert's driving.

The mean of all section compound percentages over a complete Performance and marking route is called the total mark, reflecting the general similarity of the driving performance of the student relative to the expert's driving behaviour. This means that, despite their different length and/or character, section scores are not weighted.

Shortcomings of the *Performance and Marking* system and Specifications for a new PMF system

The original PMF system (i.e., the Performance and Marking system) shows problems, ranging from minor shortcomings in the clarity of the output presentation to major flaws in the selection and calculation of appropriate performance measures. The number of specific problems that can be identified is large; it would take much space to go into each particular problem. Therefore the present chapter only discusses these shortcomings on a general level.

This discussion will follow five principles. These are of a general character such that they are also relevant for other kinds of driving simulators. In this section, the manner in which the Performance and Marking system is in disagreement with each principle will be briefly discussed. Also, a procedure for selection of (aspects of) subtasks for evaluation will be provided and a more optimal method of measurement will be indicated.

1. Objective performance measurement and explicit feedback should refer to only those subtasks that can be trained with sufficient functional validity.

The benefit of a performance measurement and feedback system (PMF) for training increases with the validity of the simulator with regard to the task. Increasing the objectivity and specificity of performance evaluations has no value if the skills that are evaluated differ from the skills needed in the operational system. It will thus be evident that using a PMF system for the training of these kind of subtasks only costs extra time. Therefore, the use of such a system should be limited to the part tasks which are simulated with sufficient validity. Hence, the development of a PMF system should start with a description of the training objectives and a task analysis, in which the task to be trained on the simulator is analyzed into its

components, or subtasks. In general the functional validity of the tracked-vehicle simulators differs for different subtasks. Subtasks that consist mainly of procedures and/or require interaction with artificial parts of the task environment generally allow for a more valid simulation than subtasks that require interaction with the natural environment (Korteling, 1990b; Korteling & Padmos, 1990). The main problems of the involved tracked-vehicle simulators concern the simulation of the normally available spatial and mechanical information about the natural environment and the degree of variation and density in the simulation of other traffic. Based on two reports (Korteling, 1990b; Korteling & Padmos, 1990) that document a task analysis and an inventory of the structural problems of the simulators, the following list of subtasks that probably will be trainable with sufficient effectiveness may be taken as a starting point:

Route driving
- driving right on straight roads
- driving left on straight roads
- stopping/braking
- shifting gears
- driving on road curves
- driving on sharp curves and at intersections
- turning on the spot

Special actions
- narrow passage ("funnel")
- "slalom" course
- vehicle clearing course ("lane change")
- parking the vehicle ("garage")
- parking on a railway wagon
- us of the short brake levers
- driving on visual signals
- driving with an image intensifier
- parking on a lowloader

Obstacle driving
- step up ("concrete block")
- sloping block
- knife edge
- small ditches (slowly)
- small ditches (quickly)
- large ditch
- cambers (normal, adverse, alternating)
- water wading

2. Objective performance measurement and explicit feedback should refer to the most critical and relevant subtasks of the driving task, while including a broad range of skills necessary for driving performance.

In order to use a PMF system as efficiently as possible, objective measurement and explicit feedback should aim at the most critical and relevant subtasks. This means that the system should not include trivial and/or overlapping subtasks. Also, the total of PMF measurements has to cover a broad range of driving skills as much as possible. The Performance and Marking system in its original form included trivial as well as overlapping subtasks. Moreover, hardly any of the special actions implied in the training of Leopard 2 and YPR-765 drivers (e.g., slalom course, vehicle clearing course) had been chosen for monitoring. In order to select key subtasks such that performance evaluations are valid and useful feedback is provided the instructors working with the simulator were consulted. Seven subtasks were qualified as trivial: turning on the spot, large ditch, small ditches (quickly), driving on visual signals, use of the short brake levers, driving with an intensified image, and water wading. Therefore, these subtasks were discarded from the list above. Primarily, these subtasks demand knowledge about simple procedures or actions in order to be well performed (Korteling and Padmos, 1990).

There is also overlap between some of the remaining subtasks. The necessary skills for driving on a straight road (keeping a good lateral position) and shifting gears (choosing the right gear/speed) are largely involved in driving on road curves such that both can be evaluated in a road course with curves. Furthermore, the step up and the sloping block are comparable subtasks that may be evaluated according to the same principles and procedures.

With respect to the special operations, narrow passage and parking the vehicle do not add much to the vehicle clearing course. In each subtask the driver has to drive between closely-separated obstacles. However, only the vehicle clearing course explicitly requires the driver to make some difficult (re)positioning operations. Also a large overlap exists between the railway wagon and the lowloader. Both tasks require the driver, guided by a marshaller, to park a YPR-765 on a transport vehicle. The lowloader is the most difficult subtask since this vehicle contains a small bump that must be taken (which also causes the marshaller to be out of sight for a moment). Therefore the railway wagon was eliminated. For PMF evaluation, the following subtasks remained on the list:

Route driving
- stopping/braking
- driving right on straight sections and on curves
- driving left on straight sections
- driving on sharp curves and at intersections

Special actions
- "slalom" course

- vehicle clearing course ("lane change")
- lowloader

Obstacle driving

- step up and sloping block
- small ditches (slowly)
- camber (normal, adverse, alternating)

3. Performance evaluation and feedback should focus on the measures that reflect the most critical aspects of subtasks.

Because different subtasks are based on different perceptual information and actions (task variables), the most critical aspects of a subtask may be different for different subtasks. For example, speed control on an YPR-765 becomes very critical when driving in sharp curves, whereas this subtask is of secondary importance on straight roads. This means that for different subtasks different critical variables are relevant to represent the quality of driving performance. This issue was not addressed in the Performance and Marking system. In this system the same broad range of variables was measured for nearly every manoeuvre. The only differentiation that has been made is the differentiation between route driving and obstacle driving. Consequently many performance measures that were presented gave no information or gave useless information concerning the subtasks involved.

Table I The selected subtasks and their critical task variables for the YPR-765.

Subtask	Critical task variable
Route driving	
Stopping/braking	lateral position
Driving right straight/curves	lateral position
Driving left/straight	lateral position
Sharp curves and intersections	lateral position
	correct speed
Special actions	
"Slalom" course	lateral position
	longitudinal speed
Vehicle clearing course	lateral position
	longitudinal speed
Lowloader	smoothness
	longitudinal speed
	following visual signals
Obstacles	
Step up and sloping block	smoothness
	longitudinal speed
Small ditches (slowly)	smoothness
	longitudinal speed
Cambers	lateral position

Based on a task analysis (Korteling, 1990b), and consultation with the instructors, the most critical (important, difficult, and time consuming) task variables were

selected for the remaining list of subtasks. It may be expected that feedback concerning these task variables is especially useful to the student. With reference to the YPR, Table I shows these critical variables for each of the selected subtasks.

4. If possible, performance measures and criteria should be defined according to objective principles, based on characteristics of the vehicle, task analysis, and formal rules for driving behaviour.

Performance criteria of the Performance and Marking system were based on the assumption that for every part of a trajectory and for every variable measured there is one optimal value, which may be produced by any expert. Apart from the variability of the expert's performance, the falseness of this assumption is demonstrated by the fact that many parts of the driving task can be performed satisfactorily using different strategies. Therefore objective and unambiguous principles should be developed in order to operationalize the measurement of driving performance. Knowledge of the vehicle and the driving task offers the most substantial opportunities for this. There are usually objective limits within which the value of variables should be kept, given the driving situation (e.g., RPM while turning on the spot: 1500-2000; speed in urban roads: < 30 mph, when turning left or right the direction indicator should be used; and when approaching the step up or the sloping block, driving speed should be decreased until one drives at walking pace and these goals should be achieved as smoothly as possible. The relevant measures and criteria may easily be implemented in a new PMF system, such that performance can be judged without the intermediary of an instructor.

Below, these kinds of absolute performance measures and criteria will be defined globally for selected subtasks for the YPR-765. A more detailed description can be found elsewhere (Korteling, 1990a; Korteling, 1991).

Stopping/braking

When an YPR-765 driver stops, he has to release the gas pedal and pull the two braking levers such that the vehicle stops in a straight course. Maintenaning a straight course while stopping is an especially difficult and important aspect of this subtask. The degree to which this is accomplished may be measured by calculating the standard deviation of the vehicle course during the time both brake levers are pulled and the vehicle's deceleration exceeds a specific value. The mean value of all measured standard deviations during the PMF route is a measure of stopping performance.

Driving right on straight sections and in curves

With respect to lateral position the student should drive always as steadily as possible on the right side of his lane and he should not drive into the verge. The degree to which this is accomplished may be measured by separately calculating the root-mean-squared (RMS) error of the vehicle relative to the right edge of the road and the total longitudinal distance over which the vehicle drives on the verge. A high RMS error reflects poor steering performance. Vehicle reference points for RMS calculations may be located at the longitudinal middle of the vehicle model.

By measuring the distance of verge driving instead of the duration or frequency, the speed as well as the time of verge driving is taken into consideration. The higher the speed and the longer the duration, the higher this index.

Driving left on straight sections

This subtask contains the same kind of measures as the prior one, except left and right are interchanged.

Sharp curves and intersections

Since gear choice is mainly determined by the radius of the curve and the width of the road, performance may be evaluated according to the criteria based on these variables. The system should then "know" rules like: when the curve radius of a road with a width of y m is between r_l and r_h m, the curve should be driven in gear position z.

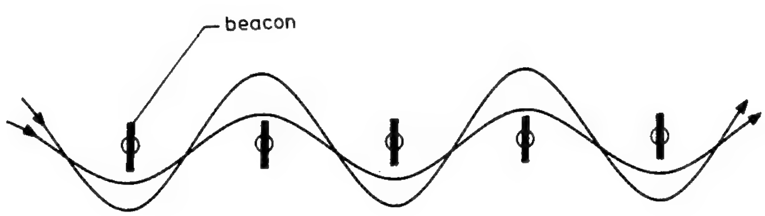


Fig. 2. Two possible manners of driving the slalom course

Slalom course

A slalom course usually consists of a number of cones in a row. The driver has to steer his vehicle in gear "1" around the beacons without hitting them. Since there are many ways to drive a slalom course correctly (Fig. 2) it is not possible to define an absolute criterion for lateral position that is more valid than the number of cones that are hit.

As a consequence of the limited field of view in the simulators used and the absence of mirrors which enable the driver to monitor his own driving behaviour, intrinsic performance feedback in this subtask is very scarce. In order to enhance performance feedback to the student, a clear audible signal in the driver's cabin should indicate the moment the vehicle hits a cone.

Also, the time taken to drive the course may be measured in order to represent the efficiency of driving performance.

Vehicle clearing course

A vehicle clearing course consists mostly of one lane change to the left and one again to the original lane (Fig. 3). The driver has to steer the vehicle as well as possible in the middle of the lanes marked out by cones. By proper gas control he also has to maintain a specific gear setting. Task performance may thus be

indicated by three absolute criteria: 1. the RMS error relative to the midline of the lanes, 2. the duration of driving in a wrong gear, and 3. driving speed. In order to enhance performance feedback to the student, a clear audible signal (see slalom course) in the driver's cabin should indicate the moment the vehicle hits a cone.

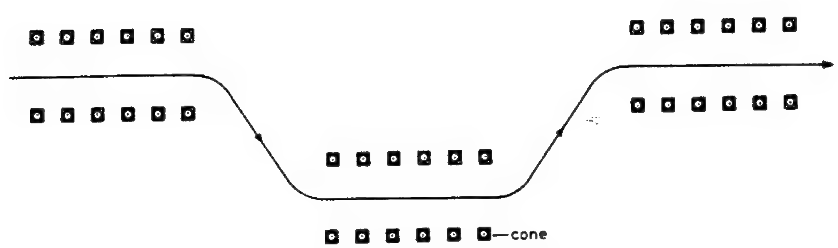


Fig. 3. Schematic representation of the vehicle clearing course.

Lowloader

A lowloader is a heavy truck designed to transport tracked-vehicles (Fig. 4). Since there is just enough space for one YPR-765 vehicle, the driver has to follow signals of a marshaller when parking his vehicle on a lowloader or when driving off. Ascending as well as descending should be performed very carefully. This may be accomplished by maintaining a low driving speed and accurate brake pedal use. When this is not appropriately done, jolts may be found in the acceleration profiles of the surge, heave, and pitch degrees of freedom.

Also, the RMS error relative to the (virtual and extended) midline of the lowloader has to be measured.

Finally the fluency, or rapidness, of driving behaviour determines the quality of task performance. Therefore mean driving speed during this subtask should also be monitored.

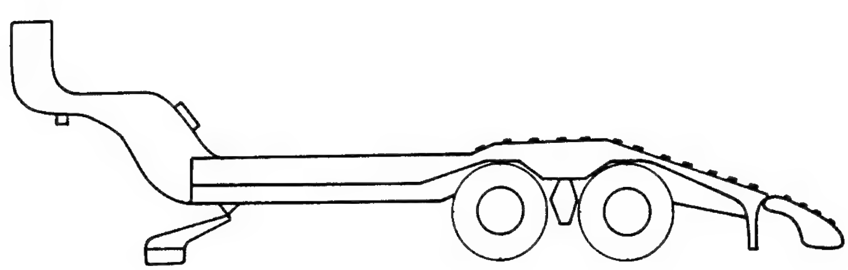


Fig. 4. Schematic representation of the lowloader.

Step up and sloping block

Performance on crossing the step up and the sloping block (concrete objects with vertical sides or steep sloping sides, respectively, including a traverses) mainly determined by the smoothness and fluency of driving. Therefore, for this subtask, the same compound smoothness-measure and speed measure may be calculated as for driving on and off the lowloader.

Slowly crossing small ditches

For crossing small ditches, smoothness and fluency of driving are also the critical performance variables. Therefore, for this subtask, the same compound smoothness and speed measure (using the same start and end points) may be calculated as for the lowloader and for crossing the step up and sloping block.

Camber (adverse, alternating)

Camber driving may include a normal camber, an adverse camber and a section with continuously changing cambers (alternating). The main problem of driving over a camber or an adverse camber is to keep the vehicle in the optimal lateral position. Therefore the RMS error relative to the right edge of the road is the best representation of task performance (see section "Driving right on straight sections and in curves"). For the alternating camber the problem is to maintain a straight and stable course by steering against continually changing lateral slopes of the road. This means that over this section just the standard deviation (deviations relative to ones own mean lateral position) should be measured.

5. Measures, scores and criteria should be easy to comprehend and implications for behavioral improvement should be clear.

With the Performance and Marking system it was often obscure what exactly was measured. For example, when the print in Fig. 1 shows mean and maximum scores on "steering" or "braking" the metrics and criteria that have been used to calculate the scores are unclear. When it is unclear which aspects of particular actions are measured, one prominent goal of a system for performance evaluation and feedback is not attained, namely: enhancing the clarity and specificity of behavioral feedback. Consequently the student still has to improve his driving performance by inefficient trial and error learning.

Secondly the prints consisted of weighted basic scores that only became meaningful after comparing them to the expert's scores and relating them to their respective weights. These requirements make the interpretation of scores and marks on the different Performance and Marking measures difficult.

The efficiency of a performance evaluation system will increase substantially when it is clear to the instructor as well as to the student which aspects of driving behaviour are measured. In addition, the scores and marks on prints should be specific and easily interpretable. Therefore two kinds of indications of the quality of a student's performance relative to the described absolute criteria may be presented. First, simple raw scores, such as the number of cones hit or the number of gear

changes. Second, transformed scores ("marks"), indicating the quality of driving behaviour according to a certain scale (such as the point system formerly used at Dutch elementary schools). Raw scores provide absolute information about the concrete consequences of a student's driving actions. Transformed scores, or marks, directly provide information concerning the level of a student's driving skills as related to the driving performances of the other students. The system can do this by relating scores to the performance of other students. This relation can easily be made when raw scores of prior students with the same training experience are saved. The most unambiguous transformed feedback then will be the presentation of percentile marks based on the scores of the students with the same level of prior training. Percentile marks, however, do not provide criterion-related information about a student's driving performance, indicating what is already learned and how performance relates to the training objectives. Therefore learning marks are necessary. A learning mark expresses the performance level of a student relative to the baseline level (mean raw score of absolute beginners) and the ultimate criterion level (average raw score of students who passed the final examination).

It would be optimal to present scores on subtasks in all three forms, raw scores, percentile marks, and learning marks.

For the three task clusters –route driving, obstacles, special actions– separate total scores have to be calculated. The most obvious cluster score is simply the mean of the relevant percentile scores. However, the subtasks within a task cluster and the measures within a subtask are not always of equal significance. This means that the scores for the different measures have to be weighted. The same applies for the three task clusters, although, for the present case, it was not considered necessary to combine these cluster scores to one total score. In consultation with the instructors working with the simulators, weights were determined such that within each task cluster the sum of the weights was 1.0 and the individual weights reflected the relative importance of the implicated measures. By adding the products of the percentile scores and their weights for all measures within a cluster, the system can compute mean scores for the three task clusters. Because weighing of raw scores will affect the interpretation of total scores, the weights have to be presented clearly on the printout.

Conclusions

The original system developed for automated performance measurement for the training of drivers on a Leopard 2 and an YPR-765 driving simulators, may be characterized by a strong engineering approach. This so-called 'Performance and Marking' (PAM) system did not take into account the human factors of performance evaluation and feedback. Therefore, this system showed many problems, ranging from minor shortcomings in the clarity of the output presentation to major flaws in the selection and calculation of appropriate performance measures.

Based on a framework of five principles, which were applied in sequence, the present paper showed how a specific new system for automated performance measurement and feedback may be developed. Fig. 5 presents a summary of the main conclusions of the former sections concerning a new PMF system for the YPR-765 and Leopard 2 driving simulators of the Royal Dutch Army. If properly implemented, this system would provide a pattern of objective grades on relevant aspects of a students driving behaviour, which is easy to comprehend. Moreover, this system would enhance the feedback to the student (knowledge of results). Apart from objective evaluation, the pattern of grades would also enable knowledge of progress and of persistent shortcomings in the students driving skills, such that the output may also be used for remedial teaching objectives (for example, when lessons are continued on the operational tracked-vehicles).

TASK	WEIGHT	VARIABLE	raw score	%	learning%
Route Driving	**0.50**				
Driving right straight/ curves	0.18	RMS lane error (cm)	cm	-% cat	%
	0.07	Distance of verge driving (m)	m	-% cat	%
Driving left straight	0.18	RMS lane error (cm)	cm	-% cat	%
	0.07	Distance of verge driving (m)	m	-% cat	%
Sharp curves and intersections	0.18	RMS lane error (cm)	cm	-% cat	%
	0.07	Distance of verge driving (m)	m	-% cat	%
	0.07	Duration in wrong gear (s)	s	-% cat	%
Stopping/braking	0.12	Lateral instability (cm)	cm	-% cat	%
	0.06	Mean deceleration (m/s²)	m/s²	-% cat	%
Obstacles	**0.25**				
Step up	0.13	Jerkiness (m/s³)	m/s³	-% cat	%
	0.03	Mean driving speed (km/h)	km/h	% cat	%
Sloping block	0.13	Jerkiness (m/s³)	m/s³	-% cat	%
	0.03	Mean driving speed (km/h)	km/h	% cat	%
Small ditches (slow)	0.26	Jerkiness (m/s³)	m/s³	-% cat	%
	0.06	Mean driving speed (km/h)	km/h	% cat	%
Normal camber	0.12	RMS lane error (cm)	cm	-% cat	%
Adverse camber	0.12	RMS lane error (cm)	cm	-% cat	%
Alternating camber	0.12	Lateral instability (cm)	cm	-% cat	%
Special Actions	**0.25**				
"Slalom" course	0.07	Number of beacons hit	n	-% cat	%
		Time needed (s)	s	-% cat	%
Vehicle clearing course	0.26	RMS lane error (cm)	cm	-% cat	%
	0.09	Duration in wrong gear (s)	s	-% cat	%
	0.09	Mean driving speed (km/h)	km/h	% cat	%
Lowloader	0.18	RMS lane error (cm)	cm	-% cat	%
	0.18	Jerkiness (m/s³)	m/s³	-% cat	%
	0.06	Mean driving speed (km/h)	km/h	% cat	%

Fig. 5. A summary of the subtasks, weights, measured variables, and performance metrics that should be included in a new PMF system.

The new PMF system does not contain criteria for examination. Criteria, or cut-off scores, provide immediate information concerning the question of whether or not a student's driving performance is sufficient with respect to specific training objectives. Based on these, it can be decided whether or not a student should be admitted to the next training phases. This kind of criterion may only be implemented after empirical investigation. Another limitation of the new PMF systems is that performance measurement is completely based on vehicle behavior. Hence insight

in the way students actually handle the controls and monitor vehicle systems and the environment is not provided. Finally, a PMF system as described forms a small part of a totally autonomous and interactive instructional system. The development of autonomous instructional systems is a very complicated matter, requiring knowledge concerning a well-defined driver model with criteria for various degrees of incorrect behavior, systems to directly register and evaluate a driver's perceptuo-motor acts, training modules offering training situations focussed at particular training objectives, a training model that is able to evaluate behavior based on which training modules are ended, repeated, or changed, and finally, multiple instruction and feedback procedures. It is clear that the development of such a system that adequately performs the majority of the instructor's task will require a giant amount of work.

References

Adams, J.A. (1979). On evaluation of training devices. *Human Factors, 21*, 711 - 720.

Adams, J.A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101*, 41 - 74.

Korteling, J.E. (1990a). *Proposal for a new performance evaluation and feedback system for the Leopard 2 driving simulator*. Report IZF 1990 A-40, TNO Institute for Perception, Soesterberg.

Korteling, J.E. (1990b). *The YPR-765 driving simulator as a training device*. Report IZF 1990 A-29 (in Dutch). TNO Institute for Perception, Soesterberg.

Korteling, J.E. and Padmos, P. (1990). *The Leopard 2 driving simulator as a training device*. Report IZF 1990 A-3 (in Dutch). TNO Institute for Perception, Soesterberg.

Korteling, J.E. (1991). *Proposal for a new performance evaluation and feedback system for the YPR-765 driving simulator*. Report IZF 1991 A-32, TNO Institute for Perception, Soesterberg.

Korteling, J.E. and Padmos, P. (1992). *Technical specifications for the PMF systems of the Leopard 2 and YPR-765 driving simulators*. Report, IZF 1992 A-20. TNO Institute for Perception, Soesterberg.

Schmidt, R.A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82*, 225 - 260.

Schmidt, R.A. (1988). *Motor control and learning: A behavioral emphasis* (2nd ed). Champaign, IL: Human Kinetics.

Behavioural adaptation to an enforcement and tutoring system –A driving simulator study

Dick de Waard, Monique van der Hulst & Karel A. Brookhuis
Traffic Research Centre
University of Groningen
The Netherlands

Abstract

Aim of the DETER (Detection, Enforcement & Tutoring for Error Reduction) project is to develop a tutoring and warning system that increases traffic safety by reducing the number of traffic violations. Detected deviations from normative behaviour trigger feedback and tutoring messages of the system. In case a tutoring message is not followed, the system has the capability of forcing behaviour adaptation by means of registration and punishment.

A prototype of this system was tested in a simulator. During the study driving parameters, mental load indicators, and subjective ratings of acceptance were collected. Two groups of drivers of different capabilities were tested, elderly and relative young drivers.

The feedback and tutoring messages were found to be successful in decreasing the number and the extent of traffic law violations. In a situation with relative high information load drivers were more likely to make the 'safe decision' when driving with the enforcement system switched on, as compared with driving without the system. However, mental effort, both reported on a subjective rating scale and measured as reduced heart rate variance, was slightly increased during tutoring.

All drivers consider the system useful, elderly drivers use the system as driver support. Younger drivers found the system less pleasant than did elderly drivers.

Introduction

Only in a minority of cases has the technical state of the vehicle or road been shown to be the principal cause of a traffic accident. Human error causes most of them (e.g., Smiley & Brookhuis, 1987); very frequently a driver's misjudgment of a traffic situation or a misperception precede a crash. Apart from inaccurate perceptions, reduced overall vigilance, as a result of sleep deprivation, alcohol or drug use, also contributes to the amount and seriousness of accidents. Given the diversity of possible traffic errors and the fact that these situations occur in a wide range of circumstances it is remarkable that most errors are covered by traffic law. A driver who misses a one-way sign and enters this road from the wrong direction

is violating a traffic rule. The same applies to missed speed signs that may lead to (unintended) speeding, an offence that has a strong link with accident seriousness (e.g., Joksch, 1993). The law is less specific regarding driver state, the exact legal criteria of e.g. drug-blood concentrations under which it is safe to continue driving, are non-existent, with the exception of alcohol. Blood-alcohol levels are strictly defined in many countries, but the Blood Alcohol Concentration level at which the law considers performance to be affected differs as much as between 0.0 and 1.0‰ in the European countries alone (see e.g. Melchers, 1994, for the BAC values in the European Union countries). Traffic law regulation regarding driving under the influence of drugs, e.g. hypnotics, is mostly very vague. In the Netherlands, for instance, the law states only that driving under the influence of drugs is not permitted if the driver can suspect that these drugs may affect driving performance. It is clear that legal issues regarding drug-blood concentrations and driving performance are complex, not only because the relationship between the two is complex, but also because of the large variety in available drugs. On the other hand, all that the law wants to prevent is driving while impaired. This impaired driving can be related to vehicle performance parameters. In fact, the police use vehicle performance (swerving or weaving of the vehicle) in selection of drivers that are suspected of driving while intoxicated. The affected vehicle performance parameters offer the opportunity for use as impairment parameters and could be watched continuously by an in-vehicle Driver Impairment Monitor (DIM). Such a DIM is a subpart of a monitoring system that is aimed for in the DRIVE (Dedicated Road Infrastructure for Vehicle safety in Europe) project DETER (Detection, Enforcement and Tutoring for Error Reduction; Brookhuis & Oude Egberink, 1992). In previous demonstration studies the relationship between driver state, as indicated by physiology, and vehicle parameters have shown that the development of a monitoring device on the basis of vehicle parameters alone was feasible (Brookhuis & De Waard, 1991, De Waard & Brookhuis, 1991). In ensuing work in the DETER project specific (steering wheel) measures were developed, and these measures have been tested and shown to be successful in the detection of reduced driver vigilance (for more details see Fairclough, 1994).

In the project the increase in safety that is to be expected as a result of error reduction by improved traffic rule compliance was also acknowledged. The effects of reduced traffic law violations on safety can be expected to be quite large, analysis of accident databases has shown that 92% of all accidents were preceded by the violation of at least one traffic law (Rothengatter, 1991). Law enforcement is one of the methods to increase traffic law compliance, but traditional policing is, just as its effect, location limited (e.g., De Waard & Rooijers, 1994). An in-vehicle offence-detection system that is continuously active would probably have a greater positive effect on law compliance. Such a system requires communication facilities to the road infrastructure (e.g., to obtain information whether overtaking is allowed or what the local speed limit is) and should compare this information with the driver's behaviour. The core aspect of the DETER enforcement and tutoring system is therefore called 'Behaviour Comparator': it compares actual behaviour with normative behaviour. In case a violation is detected by the comparator the driver is

informed about this by a tutoring message and the opportunity to correct his or her behaviour is given. If the driver ignores the message, registration of the violation is possible to which sanctioning could be coupled (the enforcement aspect). The ultimate goal of the system is to reduce driver errors by reducing the number of offences committed (Brookhuis et al., submitted).

Although it is likely that such a system will reduce violations, its effectiveness has to be assessed before introduction. Very important in this respect are false alarms, which will frustrate drivers and undermine any positive effects the system could have. Man-machine interfacing -how should the warning be presented- are also of importance. Human factors have an important role to play here, since the feedback messages and the required behavioural adaptation to them should not increase mental load. To assess driver mental load three groups of measures are available (O'Donnell & Eggemeier, 1986); task performance, subjective ratings and physiology. In driving, primary task performance could be inferred from lateral position control and steering wheel movements. Subjective assessment of mental load (see e.g., Eggemeier & Wilson, 1991) can be accomplished by ratings on scales. Finally, physiological measures can indicate mental load or effort (see Kramer, 1991, for an overview).

A group of drivers that is particularly susceptible to mental load is the elderly. More and more elderly people possess a driving licence and continue to drive (e.g., Waller, 1991). This group of people have more problems with divided-attention tasks (Brouwer et al., 1991, 1992), a type of task that increases with the introduction of more technology into the car. In addition to this, elderly are more reluctant to use technical innovations, making acceptance by this group of drivers critical (Hancock & Parasuraman, 1992). In the below reported experiment two groups of target users have driven with a prototype Behaviour Comparator in a driving simulator. Effects of the system on driver behaviour and mental workload were assessed, and ratings of acceptance were collected.

Method

task environment

In an experiment the Behaviour Comparator's functioning and its effects on driver behaviour and workload were studied using the driving simulator of the Traffic Research Centre. The Driver Impairment Monitoring module was not included in this version yet. The simulator's graphical workstation, IRIS, is a Silicon Graphics 340VGXT 'Skywriter'. Subjects completed four sessions in the simulator, each lasting about 20 minutes, and they had to drive a modified handshifted BMW 518 by original controls. Graphics were projected on a 2 by 2.5 m projection screen. Other traffic that was present in the simulated world employed hierarchically structured decision rules that are based on models of human car driving. All traffic interacted with each other and with the simulator car (for details, see Van Winsum & Van Wolffelaar, 1993). Subjects drove through built-up areas, on dual-carriageways and 'A' roads while they were guided by sampled vocal route messages. They passed traffic lights and two roundabouts. During the first and the

fourth session no feedback about violations was given, during Session 2 and 3 auditory and visual feedback was provided in case of a detected violation, the order of the two modalities being balanced across subjects. Auditory messages were presented by a digitized female voice. Maximum message duration was three seconds. The visual messages were the textual counterparts of the vocal messages. The text was printed in white, just above the horizon. This condition represents a simulation of the best possible head-up display.

The enforcement and tutoring system monitored four offences: speeding, not coming to a stop before a stop sign, red light running and entering a one-way road from the wrong direction. In case a violation was detected, feedback was provided without delay, but only in Session 2 and 3. During Session 1 and 4 violations were only registered. The speed violation messages contained a reference to the local limit, e.g., 'You are driving too fast, the current speed limit is 100'. The general instruction to the subjects was to drive the way they normally would.

To enable subjects to form an opinion about the enforcement and tutoring system they had to make a reasonable number of errors. This was accomplished by several violation enhancement scenarios that were as realistic as possible. Amongst these were: wide lanes with a relative low speed limit, other speeding cars, a complex junction, busy junctions and roundabouts, and a speed limit that took effect directly at the sign. None of these situations was unrealistic and most of them could be encountered in the real world, though probably not as near to each other as here. The route that was followed in Session 1 was slightly different from the other three sessions, in Session 1 the one-way road and a complex junction were avoided. The reason for this was that it was considered unlikely that a driver would enter the same one-way road from the wrong direction twice. In that case the driver's reaction to a tutoring message could not be assessed, because none of the drivers received these messages during the first session. The route subjects drove during the Sessions 2 to 4 were identical.

subjects

Both males and females from two age groups were recruited, elderly subjects between 60 and 75 years, and relatively young subjects, between 30 and 45 years of age. Twenty-nine subjects, of which 10 were elderly, were invited to complete the simulator test.

measures

Number and extent of speed, stop, one-way and red light violations were registered. At selected sections without curvature on the dual carriageways the lateral position on the road and steering wheel position were measured. Of both measures the SD (Standard Deviation) was calculated, both SDLP (Standard Deviation of Lateral Position) and SD Steering Wheel have been shown to be primary task parameters sensitive to task performance (e.g., Brookhuis et al., 1985, De Waard & Brookhuis, 1991).

Subjective ratings effort were collected using the BSMI (BeoordelingsSchaal Mentale Inspanning, Zijlstra & Van Doorn, 1985). This is a unidimensional scale that, if an overall demand rating is required, is to be preferred over more complex multidimensional scales (Hendy et al., 1993, Veltman & Gaillard, 1996). After each session subjects indicated on the BSMI scale the amount of effort they had invested in the driving task.

As physiological parameter heart rate was measured. From the inter-beat-intervals of the R-tops of the ECG, average heart rate was calculated and, after a spectral analysis, power in the 0.10 Hz band (i.e., 0.07 - 0.14 Hz) of the heart rate variability was calculated (for details, see Mulder, 1992). Reduced variance in the 0.10 Hz frequency band is indicative of increased mental load (e.g., Mulder, 1980, Aasman et al., 1987, Vicente et al., 1987, Brookhuis et al., 1991, De Waard, 1991, Mulder, 1992). Between the first and second session a three-minute heart rate rest baseline was registered. Another three-minute rest baseline concluded the simulator test. Subjects were instructed to remain silent during the experimental test rides because vocalization would affect their heart rate.

Before the simulator test had started, subjects had filled out rating scales regarding personal driving history and acceptance of an in-vehicle enforcement system. After the test rides they filled out the scales regarding the enforcement system again. Full details of this study can be found in a technical report to the European Union (De Waard et al., 1994).

Results

violations

The number of speed limit violations drivers could make was almost unlimited. During the first session, when no feedback about violations was given, both the young and the elderly made on average 10 speed violations. During the second and third session, when auditory and visual feedback was provided, the number of violations decreased significantly. However, when the equipment was switched off again during Session 4, the two groups of drivers diverged in behaviour. The elderly continued to make fewer speed offences while the young returned to their baseline level of Session 1 (see figure 1).

Not only did the number of speed violations decrease as a result of the enforcement and tutoring system, but the amount by which the speed limit was exceeded also decreased. A same pattern as with the number of violations was found, a gradual decrease for the elderly, and a decrease for the young during the tutoring sessions only.

While the number of speed violations was not restricted to specific locations, the other monitored behaviours were. The number of stop violations, red light and one-way violations drivers could make were limited. In figure 2 the average number of detected stop violations is shown relative to the opportunity to make a such a violation. If all stop signs were ignored, the proportion would be 100%. A similar

pattern as found with the number of speed violations was found. The enforcement system significantly reduced the number of stop violations. In addition to this, the average minimum speed measured before the stop sign also decreased.

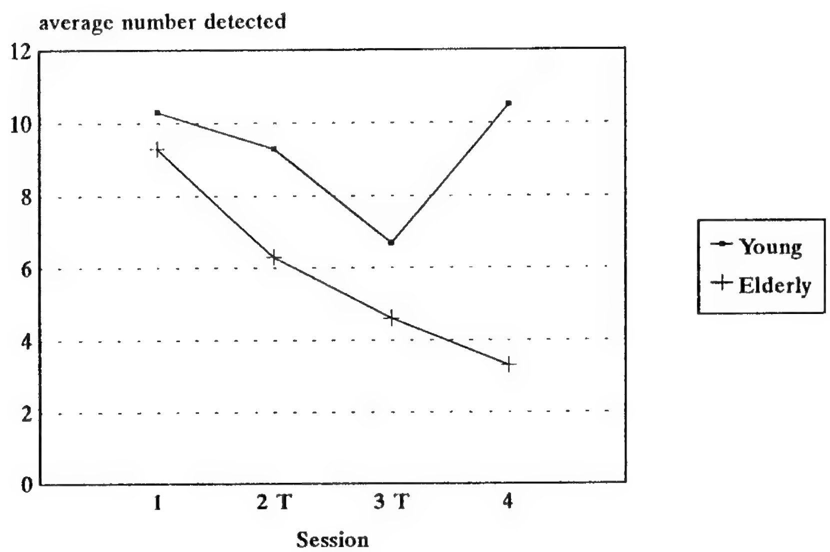


Figure 1. Average number of detected speed violations during four sessions. During the two middle sessions auditory or visual feedback was provided, which is indicated with the letter 'T' (Tutoring). Half of the subjects received auditory feedback during the second session, the other half during the third session.

As in real traffic, relatively few one-way violations were made. A total of four of these violations was detected, three during Session 2 (two violations made by elderly) and one violation, made by an elderly subject, during Session 3. During Session 1 one-way violations were not possible (the route was slightly different during this session), during the other sessions the maximum possible number of one-way violations was two.

One intersection that was controlled by traffic lights was passed during each session. The car driven by the subject set the traffic light to a 2-second phase of amber. This happened 2.2 s before expected arrival time at the junction. So, if the driver speeded up, he/she would pass amber light, while if he/she continued to drive at the same speed the light would be red. If subjects stopped for amber light, they would wait a cycle and pass green light. The amount of traffic at the intersection had been varied, two conditions were part of the experiment and were balanced across subjects: other traffic present at the intersection ('high information load') and no other traffic present ('low information load'). In the first condition subjects

were not limited in their course in any way by the other traffic if they did not stop for amber or red light. In figure 3 two bars per session are shown. Each bar represents an information-load condition. Differences in decision taking ('Go' vs. 'Stop') between information-load conditions are significant for all four sessions. A raw decision-ratio can be calculated per session by dividing the proportion of drivers that decided to drive through amber or red in the high-load condition by the proportion of drivers that made that decision in the low-load condition. If drivers are more likely to stop in a session under high information-load conditions than under low load, this ratio will be closer to zero. The values in figure 3 suggest that during the tutoring sessions (Session 2 and 3) drivers were more inclined to stop for amber light in case information load was high, compared with the non-tutoring sessions.

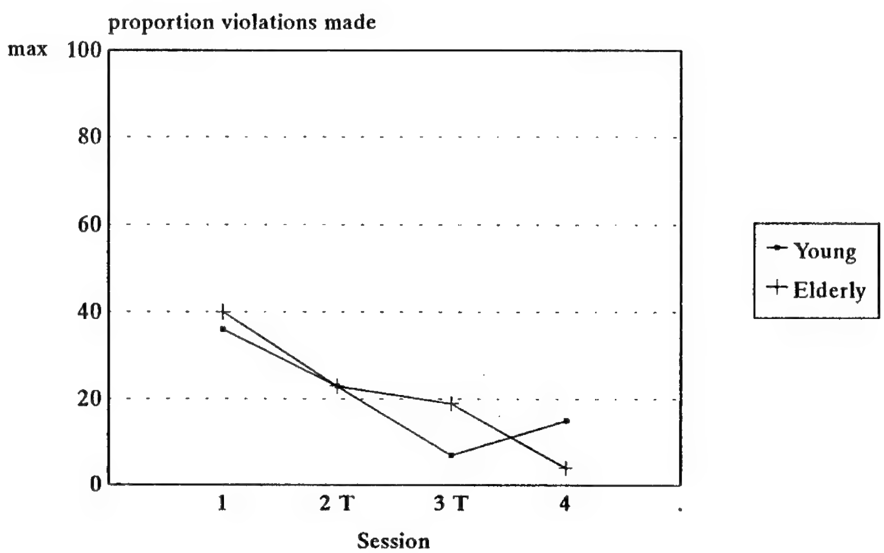


Figure 2. Detected stop violations during four sessions. During the two middle sessions auditory or visual feedback was provided, which is indicated with the letter 'T'.

mental load

Increased driver mental workload could result from the behavioural adaptation required by the enforcement and tutoring system, or as a consequence of the attention required to process the tutoring messages. Measures from three groups were taken to assess mental load; performance measures, subjective measures and physiological measures.

Performance measures were taken at straight sections of the dual carriageways (speed limit 100 km/h). The SD of lateral position (SDLP) increased over the four

sessions. The DETER enforcement device had a marginally significant effect on SDLP, during the non-tutoring sessions the SDLP was higher (34.7 cm opposed to 33.1 cm). No main effect of the feedback messages on the SD of the steering wheel measures was found.

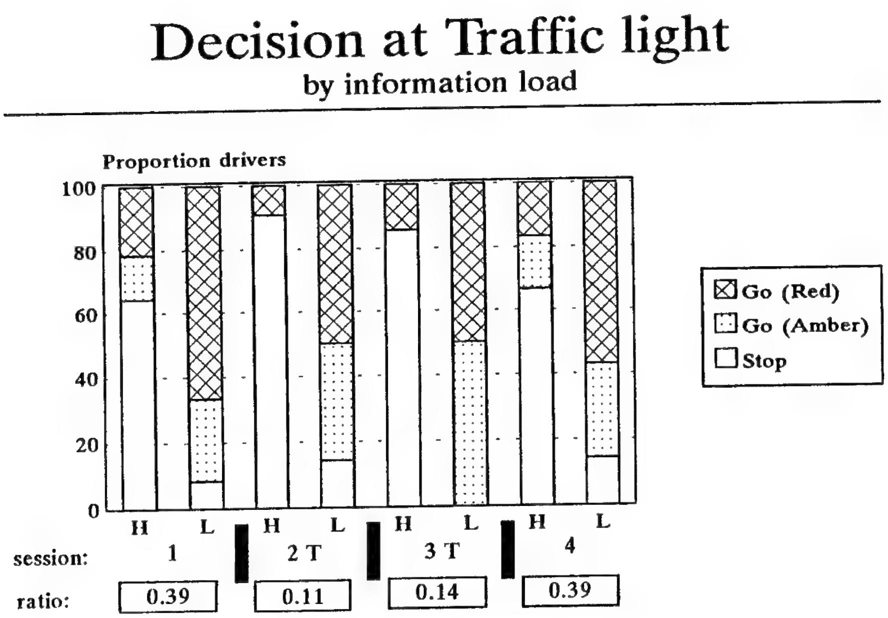


Figure 3. Proportion drivers by decision taken at amber light. Per session two conditions are depicted, a high information-load condition (H), meaning that other traffic was present at the intersection, and a low information-load condition (L) where no other traffic was present. In the H-condition subjects driving through red or amber were not limited in their course in any way by this other traffic. During the two middle sessions auditory or visual feedback was provided, which is indicated with the letter 'T'. The decision ratio, indicated below the figure, is the proportion of drivers that decided not to stop under high-load divided by the proportion of drivers that took the same decision under low visual information-load. Values closer to zero indicate increased likeliness to stop under high load conditions.

Subjective measures were collected immediately after each session. In figure 4 the average ratings on the subjective effort scale are displayed. Although the differences between sessions were small in magnitude, average effort rating during the middle two tutoring sessions was significantly elevated.

Heart rate was registered as a physiological measure. Average heart rate decreased over the four sessions, reflecting habituation to the task (e.g., De Waard & Brookhuis, 1991). In figure 5 the decrease in power in the 0.10 Hz band of heart

rate variability is shown as a percentage change relative to the rest periods. During all sessions heart rate variability is reduced, an indication of increased mental load compared to rest. During the tutoring sessions variability is reduced even further, denoting additional mental effort required by the enforcement system.

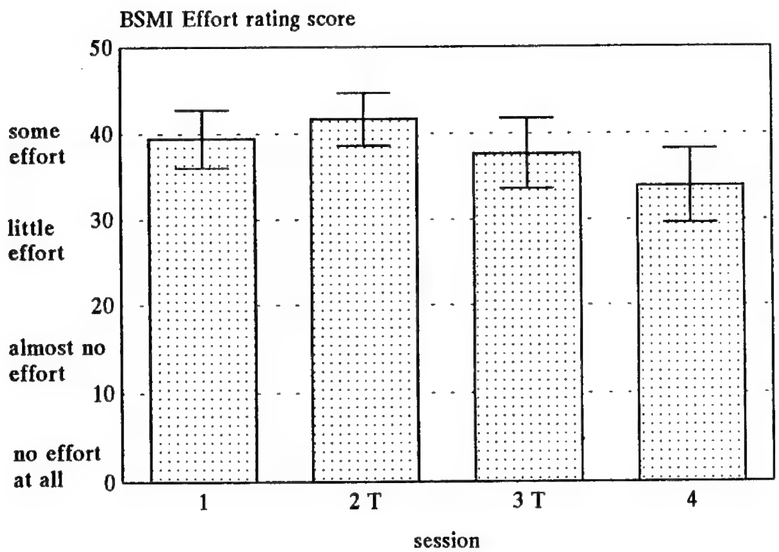


Figure 4. Average score (plus/minus standard error) on the subjective effort rating scale, BSMI, for the four sessions. During the two middle sessions auditory or visual feedback was provided, which is indicated with the letter 'T'.

Tutoring modality, i.e. auditory or visual feedback, affected none of the measures. Nor were differences in workload between the two age groups found.

acceptance

Driver opinion related to an (at that time imaginary) in-car enforcement system was asked before the first test-ride had taken place and this was compared to opinion after exposure. Subjects indicated their opinion about the system on the following nine items: 'useful', 'good', 'effective', 'assisting', 'alerting', 'pleasant', 'nice', 'pleasing' and 'desirable'. All items were 5-point scale questions with a neutral value of 0. After reliability analyses, two new sum-scales were calculated, a usefulness scale 'Practical' based on the items 'useful', 'good', 'effective', 'assisting' and 'alerting', and an affective scale 'Pleasurable', based upon the other items. Both new scales were transformed to have a range between –2 and +2.

In figure 6 the average scores on the nine items and the two sumscales are shown. From the figures it is clear that elderly drivers expected a useful system and this opinion was strengthened after exposure, while the younger drivers did not

really alter their positive opinion about the usefulness of the system. On the other hand, whereas elderly drivers had a neutral opinion about the system as being pleasurable, this belief changed in the positive sense after exposure. Young drivers were more negative about this after the test rides.

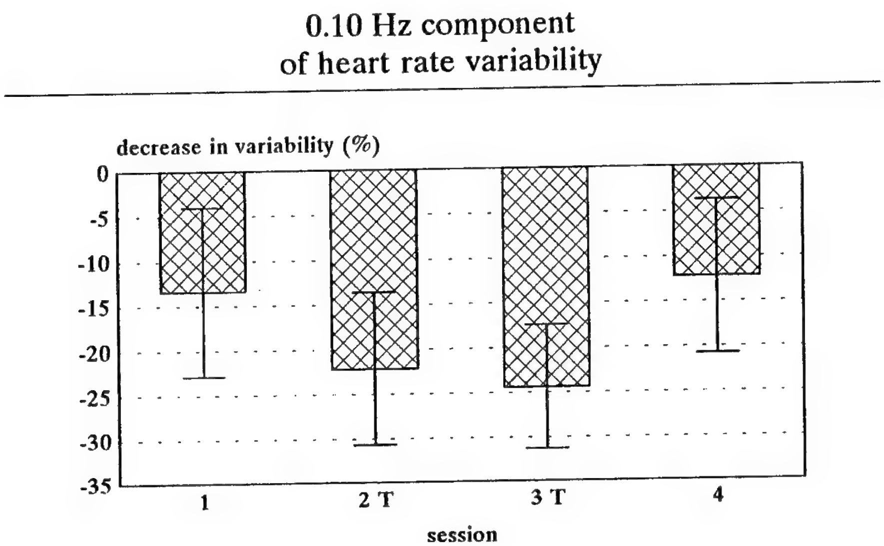


Figure 5. Power of the heart rate variability in the 0.10 Hz frequency band, compared to rest measurements. A decrease in variability indicates increased mental load. Error bars indicate the standard error of the mean. During the two middle sessions auditory or visual feedback was provided, which is indicated with the letter 'T'.

Both groups of drivers expected a large traffic-safety effect from enforcement systems, 72% expected an increase in safety while a decrease in the number of violations was expected by 83% of the drivers.

Discussion and conclusions

In the driving simulator the prototype enforcement and tutoring system was very successful in decreasing the number and extent of speed and stop violations. Moreover, in a complex situation, drivers were more likely to stop for an amber light if the system was switched on, they 'took the safe decision'. The increase measured in mental load during the tutoring sessions is a less positive result of the study. The combination of driving and receiving tutoring and feedback messages requires dual-task performance. This dual-task performance is accompanied by an increase in mental load. No differences in mental load between the two feedback modalities were found, the increase in effort in the two tutoring sessions was equal. In order to avoid violations, all signs, other traffic and the speedometer have to be closely watched. This increase in controlled attention has its costs in terms of

central resource demands (Wickens, 1984). The increase in effort was noted by the subjects, since they rated the amount of invested effort during these two sessions higher on the subjective effort scale.

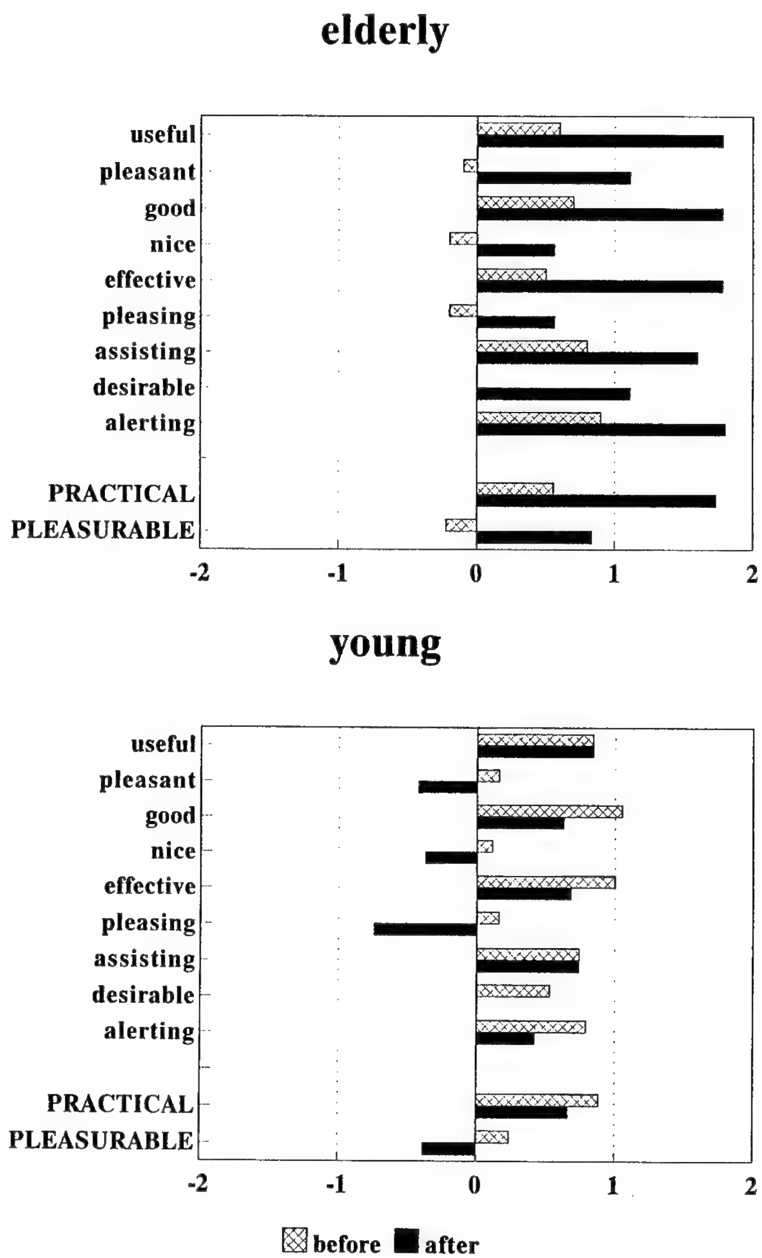


Figure 6. Opinion of the young (top) and elderly (lower) drivers about the enforcement system before and after exposure on nine individual items and the two sumscales.

Even though no additional positive effects of one of the two feedback modalities was found, auditory feedback may be preferred. Reason for this is the quality of the visual feedback messages. A 'real car' head-up display (HUD) will never be able to reach the text and contrast quality of the messages that were projected in the simulator. Moreover, voice technology is likely to be more widely available and cheaper than a HUD for in-car purposes.

In the driving simulator, situations are optimal; the vehicle's position in the simulated world is known at all times and communication with the 'road environment' is direct, without having to use special devices like tags and beacons. No false alarms were generated during the study and a minor number of offences was missed due to priority given to confidence in detection of violations (see Saaman et al., 1994). In a final, on-the-road, version of the tutoring and enforcement system hits, misses, and in particular false alarms of violations will to a large extent affect acceptance of the system. To obtain a level of acceptance that is as high as in the present simulator study the same priority should be given to a very high level of confidence in violation detection. On-the-road experimentation is definitely required to assess whether the trade-off between confidence in detection and misses of violations are acceptable, to the driver, to system performance and to the law.

The group of elderly drivers revealed some important aspects of an in-car enforcement system. The two tested groups were found to deal differently with the DETER system, both in behaviour and in opinion. The young drivers made fewer errors only if the system was switched on and even though they were convinced of the positive effects the system has on traffic safety, they disliked it. Elderly drivers on the other hand, were pleased with the system and tended to view it as a driver support system. Previous research has shown that elderly drivers miss signs more often (e.g., Brouwer et al., 1992) and it is likely that at first these drivers made errors out of unawareness. Later the tutoring messages were welcomed and drivers made use of them as road information, in particular information about speed limits.

The results in relation to the two groups of drivers have shown that introduction and acceptance of the DETER system also depends on the group the car driver belongs to. Elderly drivers may have been hesitant towards this new piece of technology at first, after some experience they changed their opinion, made use of the system as driver support and were more positive than the young about it. There is however, a risk with this reliance on the system. The DETER system is not meant as driver support, its tutoring function is restricted to the provision of warnings of imminent penalties regarding a selected list of offences, and suggested ways of avoiding these sanctions (Groeger & Chapman, 1994). If drivers rely heavily on the system to give feedback about for instance speed limits and communication with the infrastructure fails, this may lead to an unintended increase in driver errors. Younger drivers do not rely on the system in this way, they only adapt their behaviour as long as it is functioning.

References

Aasman, J., Mulder, G. & Mulder, L.J.M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors, 29*, 161 - 170.

Brookhuis, K.A., Louwerens, J.W. & O'Hanlon, J.F. (1985). The effect of several antidepressants on EEG and performance in a prolonged car driving task. In W.P. Koella, E. Rüther & H. Schulz (Eds.) *Sleep '84*, (pp. 129 - 131). Stuttgart: Gustav Fischer Verlag.

Brookhuis, K.A., De Vries, G. & De Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention, 23*, 309 - 316.

Brookhuis, K.A. & De Waard, D. (1991). The use of physiological measures in traffic research. In C. Weikert, K.A. Brookhuis and S. Ovinius (Eds). *Man in complex systems*, Proceedings of the Europe Chapter of the Human Factors Society Annual Meeting. Work Science Bulletin 7 (pp 1 - 11). Work Science Division, Department of Psychology. Lund, Sweden: Lund University.

Brookhuis, K.A. & Oude Egberink, H.J.H. (1992). *Proceedings of the first workshop on detection, tutoring & enforcement.* Haren: Traffic Research Centre, University of Groningen. Report V2009/DETER/Deliverable 1 (500A) to the commission of the European Communities.

Brookhuis, K.A., De Waard, D., Fairclough, S.H., Vallet, M. & Planque, S. (submitted). In-vehicle system for detection of violations and impairment.

Brouwer, W.H., Waterink, W., Van Wolffelaar, P.C. & Rothengatter, T. (1991). Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. *Human Factors, 33*, 573 - 582.

Brouwer, W.H., Rothengatter, T., Van Wolffelaar. P.C. (1992). Older drivers and road traffic informatics. In H. Bouma and J.A.M. Graafmans, *Studies in Health Technology and Informatics, Vol. 3, Gerontechnology* (pp. 317 - 328). Amsterdam: IOS Press.

De Waard, D. (1991). Driving behaviour on a high-accident-rate motorway in the Netherlands. In C. Weikert, K.A. Brookhuis and S. Ovinius (Eds.). *Man in complex systems.* Proceedings of the Europe Chapter of the Human Factors Society Annual Meeting. Work Science Bulletin 7 (pp. 113 - 123). Work Science Division, Department of Psychology, Lund, Sweden: Lund University.

De Waard, D. & Brookhuis, K.A. (1991). Assessing driver status: a demonstration experiment on the road. *Accident Analysis and Prevention, 23*, 297 - 307.

De Waard, D., Brookhuis, K.A., Van der Hulst, M. & Van der Laan, J.D. (1994). *Behaviour Comparator prototype test in a driving simulator.* Report V2009/DETER/Deliverable 10 (321B) to the commission of the European Union.

De Waard, D. & Rooijers, T. (1994). An experimental study to evaluate the effectiveness of different methods and intensities of law enforcement on driving speed on motorways. *Accident Analysis & Prevention, 26*, 751 - 765.

Eggemeier, F.T. & Wilson, G.F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In D.L. Damos (Ed.), *Multiple-task performance.* (pp. 217 - 278). London: Taylor & Francis.

Fairclough, S.H. (Ed.) (1994). *Driver State Monitor*. Haren: Traffic Research Centre, University of Groningen. Report V2009/DETER/Deliverable 5 (330A) to the commission of the European Union.

Groeger, J.A. & Chapman, P. (1994). *In-vehicle tutoring and enforcement system trials*. Haren: Traffic Research Centre, University of Groningen. Report V2009/DETER/Deliverable 7 (340A) to the commission of the European Union.

Hancock, P.A. & Parasuraman, R. (1992). Human Factors and safety in the design of Intelligent Vehicle-Highway Systems (IVHS). *Journal of Safety Research, 23*, 181 - 198.

Hendy, K.C., Hamilton, K.M. & Landry, L.N. (1993). Measuring subjective workload: when is one scale better than many? *Human Factors, 35*, 579 - 601.

Joksch, H.C. (1993). Velocity change and fatality risk in a crash -a rule of thumb. *Accident Analysis & Prevention, 25*, 103 - 104.

Kramer, A.F. (1991). Physiological metrics of mental workload: a review of recent progress. In D.L. Damos (Ed.), *Multiple-task performance*. (pp. 279 - 328). London: Taylor & Francis.

Melchers,W. (1994). Verkehrssicherheit in Europa. -Bestandsaufnahme-. *Polizei, Verkehr und Technik, 8*, 232 - 235.

Mulder, G. (1980). *The heart of mental effort*. Doctoral dissertation, University of Groningen.

Mulder, L.J.M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34*, 205 - 236.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman & J.P. Thomas (Eds.). *Handbook of perception and human performance. Volume II, cognitive processes and performance*. (pp 42-1 - 42-49). New York: Wiley.

Rothengatter, T. (1991). Automatic policing and information systems for increasing traffic law compliance. *Journal of applied behavior analysis, 24*, 85 - 87.

Saaman, E., Politiek, P.P. & Brookhuis, K.A. (1994). *Specification and design of InDETER-1, a prototype driving behaviour assessment system*. Haren: Traffic Research Centre, University of Groningen. Report V2009/DETER/Deliverable 9 (321A) to the commission of the European Union.

Smiley, A. & Brookhuis, K.A. (1987). Alcohol, drugs and traffic safety. In J.A. Rothengatter & R.A. de Bruin (Eds.), *Road users and traffic safety*. (pp. 83 - 105). Assen, The Netherlands: Van Gorcum.

Van Winsum, W. & Van Wolffelaar, P.C. (1993). GIDS Small World simulation. In J.A. Michon (Ed.) *Generic Intelligent Driver Support System* (pp. 175 - 192). London: Taylor & Francis.

Veltman, J.A. & Gaillard, A.W.K. (1996). Measurement of pilot workload with subjective and physiological techniques. In K.A. Brookhuis, C. Weikert, J. Moraal and D. de Waard (Eds.), Aging and Human Factors, Proceedings of the annual meeting of the Europe Chapter of the Human Factors and Ergonomics Society, November 1993, Soesterberg, The Netherlands. Haren: Traffic Research Centre, University of Groningen.

Vicente, K.J., Thornton, D.C. & Moray, N. (1987). Spectral analysis of sinus arrhythmia: a measure of mental effort. *Human Factors, 29,* 171 - 182.

Waller, P.F. (1991). The older driver. *Human Factors, 33,* 499 - 505.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and D.R. Davies (Eds.). *Varieties of attention.* New York: Academic Press.

Zijlstra, F.R.H. & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort.* Delft: Department of Philosophy and Social Sciences, Delft University of Technology.

On-the-job training for process operators as a strategy for competence achievement - A case study

Anja Johansson[1], Anders Persson[1,2] and Birgitta Wanek[3]

[1]Department of Working Environment, Lund Institute of Technology, P.O. Box 118, S-221 00 Lund, Sweden
[2]Sydkraft AB, System Operation, S-205 09 Malmö, Sweden
[3]Work Science Division, Department of Psychology, Lund University, Paradisgatan 5P, S-223 50 Lund, Sweden

Abstract

The technical development in complex and highly automated production processes affects the process operator's work and the demands on the operator. Two significant problems are difficulties in acquiring and maintaining the necessary competence and a risk of understimulation, since a stable process does not generate many active tasks, largely because of a higher degree of automation. In this case study, we were able to make a comparison between an approach to job design for solving these problems and a design of two existing operator jobs. The results suggest that this approach might be effective in practice.

Introduction

In this paper we discuss aspects of training and learning, mainly applied to process operator work. Some results from a case study will be discussed.

The role of the human operators in modern, complex systems requires a high degree of competence. The decisions they make have serious consequences, from economic, environmental and safety points of view. Bainbridge (1987) states that the higher the level of automation is, the more important the role of the human operator becomes, but at the same time it becomes harder to achieve and maintain this competence. This is explained by the fact that the operator's work tends to increasingly consist of monitoring tasks.

There are various ways of achieving and maintaining competence at work, each having it's particular advantages and disadvantages, for example formal education, simulator training and on-the-job training.

Simulator training makes it possible to train workers for situations which cannot be tested in reality, for example for reasons of safety or economy. It also gives the operator a chance to train for events which seldom occur in the real work situation. There are, however, various limitations; it is both difficult and expensive to make a realistic simulation, and still only foreseeable situations can be built into the simulator. In order to be an efficient training tool, the simulator has to be based

on extensive theoretical models. However, if these models were good enough, they could be used as a base for further automation. This leads to the irony that the unknown parts of the process, the ones that cannot be automated and thus those for which it is most important to train the operators, cannot be efficiently trained in a simulator (Brehmer, 1994).

Rasmussen (1986) defines three levels of behaviour: skill-based, rule-based and knowledge-based behaviour. Competence at the three levels is trained in different ways. In a study of ship navigation simulators, Hansen and Clemmensen (1993) found that skill-based and rule-based behaviour can be trained in simulator exercises, whereas knowledge-based learning was more limited. This is explained by the fact that unforeseen events, during which knowledge-based behaviour is trained, do not occur in these exercises, because the subjects in this study were highly-experienced navigators. Instead, knowledge-based learning took place during the briefing and debriefing sessions. As for the skill-based learning, the authors question whether the skills trained in a simulator are appropriate, since the diversity of signals available in real world situations cannot be fully represented in a simulator.

On-the-job training, which is the learning strategy that will be discussed in this paper, is by definition realistic. There are also various aspects of the operator's work that cannot be trained through formal education or simulation. An example of this is interpersonal communication; that is, knowing whom to contact in a given situation, and how to do it. However, in most cases, switching to automatic control system gives a smoother and more regular production with fewer disturbances (Brehmer, 1989). This implies that the possibilities for on-the-job training, i.e. the number of opportunities for training by direct intervention in the process, will decrease. Different factors in the work situation affect the feasibility of on-the-job training: for example, job design, organisation and management; psycho-social factors and the design of the production system.

In this paper we will present some results from a case study, carried out at a regional control center for power system control. In this study we found an example in which on-the-job learning is a good strategy for enabling process operators to achieve and maintain competence.

The general work situation for process operators

The operator's work depends, of course, on the nature of the process in question and the way in which it can be monitored and controlled. Spontaneous changes in the process are often the source of work tasks, and also problems for the operator.

The development in the process industry has changed the character of the operator's work, from a role more like that of a craftsman to a more scientific approach to process controlling. Regulating and controlling complex processes imposes many demands that are beyond the capabilities of a human operator, is one

reason why many components of process control have been automated. The development of modern control systems usually seeks a maximum of automation.

Process operator tasks

One common description of the operator's job is "to take care of everything that turns up in order to keep the process running".

Brehmer (1989) points out five categories of job tasks that give the work of the process operator its distinctive character:

- Monitoring
- Detection of deviations
- Diagnosis of deviations
- Compensation
- Optimisation

Brehmer claims that in order to understand the demands of the job of the process operator it is not sufficient to describe the different tasks. In his research he used dynamic decision making and control theory as a theoretical basis. Edwards (1962) gave the following description of dynamic decision making:

- A series of decisions is required to reach the goal. That is, to achieve and maintain control is a continuous activity requiring many decisions, each of which can only be understood in the context of the other decisions.
- The decisions are not independent. That is, later decisions are constrained by earlier decisions, and, in turn, constrain those that come after them.
- The state of the decision problem changes, both autonomously and as a consequence of the decision maker's actions.

Brehmer (1989) adds a fourth feature:

- The decisions have to be made in real time.

This approach to the task of the process operator imposes the following demands on the operator:

- to plan his tasks and interventions in the process,
- to distinguish the changes in the process arising from his own interventions from those due to spontaneous changes in the process,
- to cope with stress by planning his own tasks, which in turn requires that the operator has knowledge both about the demands imposed by alternative strategies of running the process, and of his own ability to cope with these demands,
- to have knowledge about the process dynamics and feedback delays and to be familiar with the available (process) information.

The production process and automation in the process industry

Production in a process plant is a more or less continuous flow. The purpose of a process is often to modify the physical and/or chemical qualities of a raw product or to transform one form of energy to another.

Many processes are characterised by high risk, and the consequences of human error in terms of costs to society and human life can be severe; thus the operators must often function with conflicting goals, balancing productivity and profit against safety.

Bainbridge (1987) has described some of the negative consequences that automation has had on operator work. The more automated a system is, the more seldom an operator has the opportunity to practise, but at the same time it is all the more crucial that the operator has good knowledge of the process when something untoward happens so that it becomes necessary to take over control. This has been confirmed in a study by Skorstad (1988), where he shows that some operators in a modern, highly automated, paper pulp mills intervened with the automatic process control less frequently than operators in older mills, and that these interventions also had to do with more rare and complex problems.

Berggren et al. (1986) found, in a study on process operators in a paper pulp mill, that with a higher degree of automation the monitoring task tended incresingly to dominate and a negative relationship was found between the amount of time spent on monitoring and the job commitment.

Reason (1990) points out the importance of organisational and design factors for safety in highly-automated complex systems. Referring to some of the major disasters in modern indutrial history he shows that latent errors often lie behind this kind of accidents: at the point at which the critical situation occurred the operator seldom had a chance to do anything about it because the major mistakes were made a long time ago, often by someone else (management, designer etc.). By latent errors he means errors, the consequences of which are not visible immediately, but which can lie dormant in the system for a long time until other factors occur, which in combination can breach the system's defences. This implies that work organisation and job design are important not only to achieve a good work situation for the operator, but also for the performance and safety of the system.

Common problems associated with process operator work

In addition to the cognitive problems discussed above, which are largely a result of automation, the process operator often has problems associated with his level of arousal. If all tasks are allocated to automated and computerised systems, the only task remaining for the operator is to supervise the process. If this is the case, the operator will find it hard to fulfill this task. During the nineteen-fifties, several studies of vigilance were made that showed that it is impossible, even for a highly

motivated person, to maintain effective visual vigilance towards an information source where almost nothing happens for more than half an hour.

Johansson (1989) showed that operators with monotonous monitoring tasks during uninterrupted operation experience monotony, which causes strain. She showed that this strain has to do with the fact that the work involves understimulation, which arises when almost no disruptions occur; this is the situation faced by many operators in monitoring tasks. Johansson states that there are strain-reactions to both understimulation and overstimulation, and that these are related to stimulation levels as a U-shaped curve. This means that the strain-reaction appears with both under- and overstimulation but not at a moderate level of stimulation. The relation between performance, in information processing and problem solving, and stimulation is the reverse, an inverted U-shaped curve. At both high and low degrees of stimulation we find lower performance, but at moderate levels of stimulation higher performance is found.

Earlier research by Karasek (1981) has revealed that the most severe symptoms of strain arise from a combination of a low degree of self-determination and high work demands. This combination is often the case for many operators with a traditional work organisation and a highly automated process involving mostly monitoring tasks. The health consequences of perceived stress depend on the resources that the individual has at his/her disposal in order to cope with the situation. Access to resources has a positive effect on health at all levels of strain. Such resources can range from technical and informational support, the individual's competence, to social support.

An approach to work organisation and job design in process control

Considering on-the-job training as an alternative or complement to simulator training or education, aspects of training are important for the design of the employees' work situation. A holistic view of the design of the entire production system is necessary to satisfy the psycho-social working environment, efficiency in production, safety, etc., in addition to training aspects. To summarise, the most critical problems in operator work arising from a higher degree of automation and computerisation in process control are:

- On the one hand it might be more difficult for the operator to acquire and maintain the necessary skills and level of competence. On the other hand, modern complex processes and control systems impose higher demands on the human operator, e.g. in problem solving in unforeseen situations.
- There is a risk of understimulation since a stable process does not generate many active tasks for the operator, and this results in a low level of arousal, and vigilance problems for the operator. The infrequently-occurring disturbances in the process may suddenly raise the operator's level of arousal and then cause a high level of strain.

We suggest two complementary ways of redesigning process operator work:

- Through job enlargement, new tasks, usually tasks surrounding the existing control tasks, are added to the operator's work. Primarily, we find these new tasks in areas such as quality control, and preventive maintenance of the process equipment. Other similar tasks might be fault diagnosis or error recovery. It has been suggested that the operators should take over some of the foremen's tasks. By the addition of such tasks the operator is given an opportunity to get an overview and a wider understanding of the process.
- A deepening and specialisation of the operator's role, aimed at combining experience based knowledge with theoretical knowledge, might be a radical approach according to Olsson and Brehmer (1990). The operator can, for example, take part in the development of models of process sections or an entire process. The task of process control can be expanded to include on-line control of parameters in product quality, environmental influence, production economy, etc.

Olsson (1988) presents an approach to job design and work organisation in which new tasks are integrated into the job of the process operators, aimed at decreased job stress, improved learning and mental stimulation. This approach might both enlarge and deepen the operator's role. New tasks can fill in the periods of otherwise passive monitoring, and these tasks allocated to the operators should be of a more strategic nature: "Primarily, these tasks shall be aimed at retaining and improving their skills and qualifications and their acquaintance with the process and the plant. Such tasks can be found in the surrounding functions of process operation, i.e. in mechanical and electronic maintenance, in production planning and management, in quality control and in system development". Olsson (1988) suggests that a team of operators, perhaps with support from different specialists, should have the responsibility of the entire process operation. The operation team should manage its internal labour division and, by means of a job rotation system, all operators will participate and learn all tasks successively. Exactly what additional tasks and functions should be assigned to the operators normally depends on factors such as the type of process, the organisation structure, and the education of the operators. To create a work situation in which the operators can acquire a continuous learning from their tasks, Olsson (1988) proposes the following categories of tasks:

- *Preventive maintenance* is closely related to the primary function of the operator's work - to keep the process running without disruption. Many maintenance tasks, allocated to maintenance staff without requiring their special skills, could be accomplished by process operators. By performing tasks such as plant inspections, replacement of worn-out components, tightening leakages etc., the operators would be able to gain an up-to-date acquaintance with the plant and its subsystems and components.
- *Production planning functions*, on a short or medium time scale, should be carried out by operators. This will improve their understanding of different production conditions, marketing and competition problems.

- The work of the operators is of the highest importance for product quality and it would therefore be natural to give them *quality control tasks*. Feedback of quality data to the human operator can establish awareness and responsibility at a suitable organisational level at which deviations in product quality can normally be most easily corrected.
- The operators' experience from process operation should be used in *system development* (i.e., by formulating specifications and criteria) in a continuous updating and redesign of the control and production systems. The operators can, for example, design VDU displays that support monitoring or decision-making in different planning or control tasks.
- The *rationalisation process* is usually conducted by the management, but the operators can also play an important role in the everyday rationalisation process by identifying non-optimal functions. For rationalisation, as well as for training of new operators, operators should document and analyse their work procedures.
- By making *process analyses* the operators can acquire a theoretical knowledge of the process. Normally a great number of process variables are recorded and stored in a database in a computerised control system, but much of these data is not used today. If the operators had adequate equipment, e.g. PCs and suitable programs for analysis, and were trained for the task, they could make analyses of the data aimed at improving control strategies, quality control, energy and material savings etc.

The idea behind this job design is to establish an organisation in which the operators can learn continuously from their tasks and an effective production is supported by the operators' awareness and involvement in the goals of the organisation. The operation team should be given extended responsibility and must consequently have the authority to order assistance from supporting departments, for example functions such as system development or maintenance.

The case study - a regional control center for power system control

Nation-wide, and even continent-wide, integrated power systems are among the largest and most complex systems created by man. The operator's tasks in power systems cover all the categories and aspects described in the earlier section. It has been observed that even in the operation of highly-automated power networks, the need for human intervention is still great (Bibby et al., 1975). In a survey of thirteen control centers in the USA, Williams et al. (1980) observed that the operator's job makes many demands on the operator in terms of technical skill, technical knowledge, interpersonal relations, adjustments to shift work, and high, sometimes critical workloads.

Sydkraft's control center

The Electricity Business Sector in the Sydkraft Group sells and distributes electricity, mainly in southern Sweden. In Sweden there are some ten major

generating companies, of which Vattenfall is the largest, with a market share about 50%, and Sydkraft the second largest, with a market share close to 25%. Within the regions covered by Sydkraft's business sector there are in all about 1 200 000 customers and an annual sales volume of some 31 TWh electrical energy. To optimise electricity production, Nordic power companies are engaged in a co-production energy exchange system.

The transmission and distribution of electricity is conducted through wholly-owned and leased lines. A large proportion of the electricity produced at Sydkraft's hydro-power plants in northern Sweden is transmitted over the main grid for 400 and 220 kV. Power from the main grid is also transmitted through a number of regional systems, which are monitored and operated by a number of control centers.

System Operation

The department of System Operation has its control center located in Malmö in southern Sweden. The System Operation Department fulfills two main functions:

- Dispatch and Generation control: with an overall picture of southern Sweden's power requirement, and main responsibility for electric generation, it directs and co-ordinates production, balances supply and demand, buys and sells power, and is in constant touch with Sydkraft's own power stations and other energy producers in Sweden and the other Nordic countries.
- Grid Management: it has responsibility for southern Sweden's share of the grid. Along with the operational centers it has a comprehensive picture of the grid system in southern Sweden. Interruptions and disturbances to the supply have to be rectified rapidly and effectively.

Each of the two functions is managed by one operator. At the time of our study there were a total of eight operators in each operator function. Both operators work in the same control room, which is manned around the clock. The operators spend about half their working time in the control room. During the other half they carry out office work, mainly doing engineering tasks.

The most important tasks that the process implies for the Dispatch and Generation-control operator are:

- Prediction of the power load, i.e. the consumption of electrical energy, from one hour up to 24 hours ahead.

- Planning which productions units (power stations), to use on each particular occasion. Since some power stations have start-up times ranging from several hours to days, their use must be planned in advance. The planning of an optimal utilisation of hydro-power is also a complex problem.

- Switching off some parts (electric heaters) of the power load at peaks, for reasons of economy or reliability.

- Communicating with other producers to investigate the possibilities of buying and selling energy to optimise the financial outcome.
- Monitoring the balance of supply and demand, and certain other crucial parameters in the power system.
- Quickly making a new plan for the supply in case of a disturbance, for example in case of a breakdown of a power station.

The Dispatch operator normally has a rather continuous flow of real-time tasks to handle in order to keep the process in a satisfying condition.

The most important tasks that the process imposes on the Grid Management operator are:

- Monitoring the state of the grid between the 400 kV and 130 kV.
- Monitoring the execution of service orders. These orders are procedures for performing a planned interruption in a part of the grid in order to maintain the equipment.
- Updating the plan for interruptions in the grid. This plan can be changed in case of unforeseen complications, for example, a thunderstorm.
- Supervision and co-ordination of remedial actions in case of disturbances.
- Remote-controlled start-up of gas turbines when there is an sudden need for additional production.

In addition to these on-line tasks, the operator writes the switching instructions that are to be executed in the near future; normally within a couple of days. After a disturbance he has to make an analysis and write a report on the event. And when there is no other task to be performed the operator should carry out his ordinary office work. The Grid operator seldom intervenes in the process directly, or indirectly via directives to operation centers, and if he does so, it is usually to rectify some disturbance.

Apart from the different operator functions, we have found that there are many factors in the work situation that are common to both groups of operators. We found the following similarities between the two groups:

- They work in the same control room. An inspection confirmed that the physical environment is equivalent for the two operators' work places.
- They use basically the same information technology.
- They have the same shift work system. This system means that, besides the work in the control room, the operators have part-time office work; this work can be described as engineering and investigatory.
- They have the same management and personnel policy and are close organisationally.

- There are close similarities between the groups with regard to education and age-distribution (see table 1). Furthermore, both groups were exclusively male at the time of the study.

Differences between the two groups, which we believe are important, and which we cannot systematically control are:

- They have different foremen.
- Different personalities, which might have influenced the choice of profession.
- The tasks in the office work vary between the groups and between individuals.

Good psycho-social status for both operator groups

Results from the case study have been used for the test of a hypothesis concerning the two operator groups (Wanek et al., 1994). We assume that the most important difference in job design for the two operator functions is their different tasks in the control room. On the basis of theory and findings on psychosocial status and level of arousal for active versus passive operators (Johanson, 1991, Berggren et al., 1986), together with our observations of the tasks that the process implies for the two operator functions, the following hypothesis was stated.

The tasks (mainly interactive control tasks) for the operator in the Dispatch and Generation control function create an active process operator job, while the tasks (mainly monitoring tasks) for the operator in the Grid Management function create a passive process-operator job. Three methods were used for the test, namely a questionnaire, constructed from a model of psycho social work factors, a structured interview containing 66 questions on the work, and a subjective assessment of the level of arousal in different work situations. The results implied that the hypothesis should be rejected, since none of the differences that were expected from a comparison between active and passive operators were found. A comparison with a reference group of 110 process operators showed that the two groups in our study have on the whole good psycho-social status. We concluded that none of the subjects in our groups had the kind of psycho-social status that can be expected from a passive operator. The results show that the Grid operator's work situation is as satisfactory as the Dispatch operator's work situation. This is surprising, in view f our description of the different on-line tasks that the process implies for the two groups. It is primarily the results for the group of the Grid operators that are unexpected, and better than expected.

Case studies at other power companies

In a comparison with the results from the survey by Williams et al. (1980) referred to above, we find that the operators at Sydkraft's control center are more independent in the sense that there is no senior supervisor in charge who has an overall system responsibility, i.e. a third operator function, as there is in many of the control centers that have been studied. The foremen's role at Sydkraft was

focused on planning the manning of the control room and did not interfere directly with the operators' work.

Fabricius (1991) studied the operator's work at Vattenfall, the largest power company in Sweden. Many of the traditional problems in process operator work were found in the work situation at Vattenfall. The operator's work has over time become impoverished owing to the automation of the process. Tasks such as planning and optimisation have to a large extent been automated. We found that job design for the operators implies passiveness and limits the development of competence. The division of work tasks into separate process sections, instead of supporting greater task identity (see Hackman and Oldham, 1991), is also an inhibiting factor. Sydkraft's organisation for system operation is also different from the one at Vattenfall, the latter being more hierarchic. Sydkraft has had a more consistent strategy for improving competence and efficiency at their operational control centers, and has also had achieved higher competence among the personnel in the functions for Grid Management as well as for the Dispatch and Generation-control.

Methods

All subjects were informed in advance about the study and the areas to be studied: the control room, the computer systems for process control and the operator's work situation. They were also told that participation was voluntary and that the objective of the study was to obtain comprehensive information about their work. This information was also meant to be a basis for a development project aiming to give the operators improved tools for their work.

Our methods were mainly structured interviews. During the time when the design of the study was being worked out –about one month– we also observed in the control room. The study was carried out in less than two months during the winter 1992/1993.

Subjects

Table 1. Subjects in the study

	Grid Management	Dispatch and Gen.control
Average age	39	40
Age (min-max)	26-42	34-46
Average years in occupation	5.5	9.3
Years in occupation (min-max)	0-10	1-18
Average years as employee	15	16
Years as employee (min-max)	4-20	12-22
Number of technicians	7	7
Number of engineers	1	1

All 16 operators participated in the interview and the questionnaire. Eight of them were working in the Dispatch and Generation-control function and eight in the Grid Management function.

Next to their work in the control room, all the subjects had part-time office work that can be described as engineering and investigatory work. All the subjects were men.

The interview

The structured interview, comprising 66 questions, was designed after we had spent some time participating in the control room in order to get a fair understanding of the operators' work situation, and to find out which question areas would be relevant for the interview. We wanted to get a description of the operators' various work situations. We conducted pilot interviews with two former operators, one Grid operator and one Dispatch operator, to find out if we had missed any important aspects of the operators' work situation, and to find out whether any of the questions were unclear. After that we made some small adjustments.

The interview questions, mainly concerning the control room work, were categorised under five headings:

- Work tasks, primarily the work tasks in the control room.
- Control systems and information systems
- Work organisation
- Training and learning
- Social relations

It took about two to four hours to accomplish each interview. The interviews were conducted by three researchers, of whom at least two participated in each interview. With the subject's permission the interview was tape-recorded and then transcribed. All the separate answers for each question were collated.

Results

As mentioned in the methods section, the interviews contained 66 questions categorised under five different headings. In this limited space it would be impossible to cover all the information that came up in the interviews, but here is a summary of the most relevant answers:

Work tasks

A question about resources available to fulfill the goals defined for the operators' duties was answered with a great variety of answers. Among the operators in the Grid Management function, no single resource was mentioned by more than 3 of the 8 subjects. There were 3 such resources in all, one of them being "competent and experienced operators". The operators in Dispatch and Generation-control

mentioned mostly technical and information resources, for example telephone contacts with co-operating power companies. Personal resources were mentioned by half the subjects, who stressed the importance of competence, knowledge, motivation and sufficient authority for the operators.

We asked the operators about their tasks during their office work, and also whether these tasks were useful for their work in the control room. All subjects answered that most of the office tasks had a connection to the control room activities, and that these tasks gave them useful knowledge about various aspects of the process; for example knowledge about the structure of the power system and updated information about the process state.

The subjects were asked in what situations they experience insecurity or strain, and what can be done about this. The answers were different, depending on how experienced the operator was; the more experienced the operator was, the more seldom such situations occurred. About half of the operators, from both functions, mentioned that competence and experience is crucial in stressful and otherwise critical situations. Various suggestions came up about what kind of support, both human and technical, could be useful in those situations. Regarding human support, they were quite satisfied with the present situation. Regarding technical support, five of the operators in the Grid Management function suggested some kind of on-line simulator for testing the consequences of their actions. Of the operators in Dispatch and Generation-control, three suggested that a simulator for training purposes would be useful for training disturbances and other difficult situations.

Preparedness for disruption was discussed. The subjects were asked what it is that makes it possible to deal with disruptions. Among many factors competence, knowledge and experience of the operators were stressed as important by some of the operators in Dispatch and Generation-control.

We asked the operators in Grid Management what kind of knowledge and competence is achieved by writing so-called switching-instructions, and what the consequences would be if this task were removed from the operator's job. All operators answered that the writing of switching-instructions gives them knowledge and skills valuable for their tasks. All thought that removing this task from the operator job would mean a great loss of process knowledge and general competence, and also that it would be impossible for the operator to write a certain type of simplified switching instructions for emergency situations. Some of them also thought that the quality of the switching-instructions would be diminished if someone other than an operator wrote them. Most of the operators had a positive attitude towards computer support for the writing of switching instructions, but they stressed that it was not a good idea to entirely automate this task.

Training and learning

All subjects started their operator training sitting beside an experienced operator and observing him. They gradually took over more of the tasks, and were given more responsibility. All of them were pleased with this "learning by doing" system, but several subjects thought that a more structured training plan would be welcome. Simulator training was also mentioned as an element that could be useful for operator trainees.

The issue of practical experience vs. theoretical knowledge was raised. Most subjects were of the opinion that practical experience is more important than school education. It was mentioned that theoretical education is necessary, since the operator training period would otherwise have to be too long. Personal qualities, such as common-sense and being a competitive kind of person, were stressed as factors that are important for a person's potential for becoming a successful and skilled Dispatch and Generation-control operator. In both functions, knowledge of foreign languages was considered an advantage. Generally, the senior high school engineering diploma was considered an appropriate minimum level of school education for operators, combined with some experience from work in other parts of the organisation.

We asked the subjects to give suggestions about how the operator's work could be designed to achieve competence and training at work. An important factor, mentioned by several operators, is analysing disruptions afterwards and discussing them with other operators. Simulator training and continuous education for operators after the trainee period were mentioned. Most of the subjects said that it is very much up to the individual how much he learns, but that learning comes as a natural part of the job.

When asked what they thought about the competence within Sydkraft compared to other similar power companies in Sweden, most operators thought that Sydkraft's competence is at least as good as that of the other companies. One of them thought that another company may have better formal competence, but Sydkraft had better operational competence. The special strengths of Sydkraft that were mentioned were the job design for the process operators, that gives them a high degree of responsibility and authority, but also wide competence and co-operation between the Grid Management and the Dispatch and Generation-control offices in the same control room.

Analysis

The results from the interview suggest that for both Grid operators and Dispatch operators, competence is considered one of the main resources for the operator, both in normal operation and in handling disruptions and stressful situations. We also found that the operators generally felt that they possess this competence, and that experience was considered one of the most important factors for achieving it.

This indicates that the operators, in both functions, get continuous on-the-job training through their daily activities.

In a previous section we have described how the tasks generated by the process are likely to involve an active operator role in the Dispatch function and a passive role in the Grid Management function. We also refer to earlier results, which show that the psycho-social status of the Grid Management operator is better than expected from the assumption about a passive operator role, and that many of the traditional problems described above have been avoided. The results from our study suggest that this is also the case for competence achievement at work.

Categories of work tasks

In order to explain the results, we will here categorise the additional tasks for the operators in the control room, some of them described below, according to Olsson's approach to job design for process operators, described earlier.

- *Preventive maintenance*: Operators in both functions carry out maintenance of the control and information system as a part of their office work tasks. They also maintain the instruction material, by rewriting old instructions and writing new instructions for the operation, made necessary by changes in the power system.
- *Production planning*: In both operator functions, production planning is an essential part of the everyday control room work. In the Dispatch and Generation-control function this includes planning of the power supply by unit commitment, optimisation of hydro power production and buying and selling power. The Grid Management operators plan the maintenance work on the grid on a medium and short term basis.
- *Quality control*: For the Grid operators, recording statistics of disruptions is a task that fits into this category. Some of the operators have this task in their office work. One of the tasks in the office work for the Dispatch operators is to follow up statistics on the quality of the prognoses, e.g., of the power load on the grid.
- *Systems development*: The development of the computer systems for control room work is to some extent carried out by the operators themselves. Sometimes this is done by project groups in which the operators are represented.
- *Rationalisation*: Measures for rationalisation are sometimes proposed, always discussed, and often modified by the operators before they are implemented.
- *Process analyses*: Sometimes, special investigations need to be made, for example an analysis of what consequences the design of new equipment can have on the operator's work. This kind of task is present in the office work for both operator functions. The Grid operators also produce background material for decisions about sizeing the power system as part of their office work, and in the control room they make reports after each disruption, analysing the events.

Job design for Grid Management operators

In order to explain the results, a more detailed description of the Grid operator's work situation and job design is in place here. The additional tasks can be divided into two main categories:

- Office work; that is, tasks to be carried out when there is no other task at hand, varies for the individual operators. However, in the interview all the operators stated that these tasks were related in some respect to the control room work.
- Off-line tasks, which are not included in the office work. These tasks, which often have a close relation to the on-line tasks are, at other power companies, often carried out at departments separate from the control room. The most important of these tasks is the writing of switching instructions. Another of these tasks that we will discuss is the reporting of disturbances.

Using the answers from the interview and further discussions with a few operators as a base we will, in what follows, discuss what effects some of the additional tasks have upon the operators' competence and knowledge in various areas. Office work includes various tasks; these range from statistical analyses of error rates and reliability in the grid to the maintenance and updating of computer systems, simulations and analyses of the power system, administration of training and teaching operators at the control centers. By doing these tasks the operators are able to acquire a deeper theoretical process knowledge, and knowledge about parameters such as security, reliability, environmental influence, financial issues etc.

The Grid operator who is in charge when a disruption occurs must make an analysis afterwards and write a report on the event. This responsibility motivates the operator to find out immediately as much as possible about the disruption and document the relevant information. Thorough analysis and reporting of such events is a good foundation for gaining knowledge from experience. Such knowledge, gained from personal experience or the experience of others, will be of great help, for instance in prioritising different tasks and rectifying actions, in cases of future disturbances.

The writing of switching-instructions, which is basically a planning task, is the individual off-line task that in our opinion occupies most time for the operator. According to the answers from the interview, it is also considered to be the most important one, during normal process operation.

The switching-instruction is a sequence of actions to be performed when there is a planned interruption in the grid. These actions are carried out by means of remote control from operation centers or directly in substations, switching yards and on the lines by maintenance personnel. In the interview all the operators emphasised that this task provides valuable knowledge for the on-line operation of the grid, and if the task were allocated to another department it would eventually

result in a severe loss of competence of the operators. The areas in which the switching-instruction task provides knowledge or skill are:

- Knowledge about the topology and geography of substations, switching yards and lines.
- The rules and logic for switching
- Communication paths and personal relations with people in different parts of the organisation.
- Knowledge about the present switching state of the grid.

All of these areas are important and necessary in the handling of disturbances and other critical situations. In such cases it is necessary to be familiar with the structure of the grid, to have an updated knowledge of the present switching state, and to be familiar with switching for corrective purposes. In order to execute these actions it is usually necessary to contact the same people as in the case of a switching-instruction.

We conclude that the competence and the skill the operator can achieve from working with switching-instructions is of great importance for the ability to cope with the stressful and time-critical work during disturbances. It is likely that the operators' competence, achieved by on-the-job training, will help to minimise the risk of human errors in the work of rectifying disturbances and in other unforeseen situations.

Conclusions and discussion

We can state that the job design for the process operators in our case study differs from that at other similar power companies in the following significant respects:

- The organisation for the control room work is horizontal, with fewer hierarchic levels.
- For the Grid operators there are additional off-line tasks in the control room. In other power companies these tasks are often carried out by personnel at separate departments.
- The operators spend about half their working time in the control room, during the other half they carry out office work. In other companies, in contrast, this kind of job is often designed as a pure operator shift-work.

From the answers in the interview, and from our on-site observations, we conclude that the additional off-line tasks for the Grid operators, not necessary for the on-line control of the process, have positive effects in two aspects:

- These tasks give the operator a higher level of arousal during periods when there is otherwise a risk of understimulation (Wanek et al., 1994), as a stable process does not generate tasks other than monitoring. The operator will probably find it easier to cope with disturbances in the process, which suddenly increase the operator's arousal and often cause strain.

- In addition to mental stimulation from these tasks the operators have the benefit of continuous learning; retaining and improving their skills and qualifications, and improving their acquaintance with the process and the grid.

In a comparison with the suggested approach (described above) to a better work situation for operators, we find that our case has some differences from what we understand were Olsson's prerequisites:

- The process in our case is not contained in a plant, like a traditional process industry.
- The operators in our case have a higher level of education than usual a in process industry, which perhaps makes it easier to add additional engineering tasks.

However, our case has substantial similarities with the suggested approach with regard to work tasks and other aspects in job design:

- There are tasks that correspond to practically all suggested categories.
- There is a high degree of self-determination for the operators.

We therefore find that our case study gives substantial support to Olsson's approach to job design for continuous learning for process operators. For the reasons mentioned above it is, however, not clear what effects applying this kind of job design to less qualified operator work would have. This is an issue that needs further study.

References

Bainbridge, L. (1987). Ironies of Automation. In J. Rasmussen, K. Duncan and J. Leplat (Eds.) *New Technology and Human Error* (pp 271 - 283). New York: John Wiley & Sons.

Berggren, T., Henriksson, J., & Holmgård, G. (1986). *Arbetsmiljön för processoperatörer i traditionell och datoriserad massatillverkning*. Arbetarskyddsfonden.

Bibby, K. S., Margulies, F., Rijnsdorp, J. E. and Whiters, R. M. J. (1975). *Man's role in control systems*. IFAC 6th Triennal world Congress: Plenary paper P.4.

Brehmer, B. (1989). *Operatörer och styrsystemet —vad kan vi lära oss från beteendevetenskaperna?* Conference paper January 23-25, CMD, Uppsala.

Brehmer, B. (1993). *Mänsklig styrning av komplexa system*. MDA-rapport 1992:13. Uppsala

Brehmer, B. (1994). *Personal communication*. Department of Psychology, Uppsala University, Sweden.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors, 4*, 59 - 73.

Fabricius, G. (1991). *Övergripande analys av funktioner för driftövervakning - Etapp 1*. Vattenfall, Stockholm.

Hackman & Oldham, (1991). A new strategy for job enrichment. In D. Kolb, I. Rubin and J. Osland J. *The organizational behavior reader.* New York: Prentice Hall.

Hansen, J. P. and Clemmensen, T. (1993). *Cognitive aspects of learning and cooperation in simulated ship manoeuvering.* In: CSAPC '93, Fredriksborg, Danmark August, 25-27 1993, pp 374 - 402.

Johansson, G. (1989). Job demands and stress reactions in repetitive and uneventful monotony at work. *Int. J .Health Services, 19,* 365 - 377.

Johanson, G. (1991). Stress i arbetslivet. In: L. Lennerlöf (Ed.) *Människan i arbetslivet* (pp 122-139). Stockholm: Allmäna Förlaget.

Karasek, R. (1981). Job socialization and job strain: the implications of two related psychosocial mechanisms for job design. In G. Gardell and G. Johansson (Eds.), *Working life.* New York: John Wiley & Sons .

Olsson, G. (1988). *Job design in complex man-machine systems.* In the Third IFAC Conference, Oulu, Finland. June, 14-16, 1988.

Olsson, G. and Brehmer, B. (1990). *Control Systems in Process Industry.* In Proceedings of 6th International Symposium on Work Psychology, Dresden, March 27-29, 1990.

Rasmussen, J. (1986). *Information Processing and human-machine interaction: An approach to cognitive engineering.* New York: North-Holland.

Reason, J. (1990). *Human Error.* Cambridge: Cambridge University Press.

Skorstad, E. (1988). Technology and overall control: an example from the process industry. In I.V. de Keyser, T. Qvale, B. Wilpert & S.A. Ruiz Quintanilla (Eds.), *The meaning of work and technological options.* Chichester, Wiley.

Wanek, B., Persson, A. and Johansson, A. (1994). *Job design for process operators - a case study.* In Proceedings from Learning and Research in Working Life, Lund, Sweden, June 20-22, 1994

Williams, A. R., Seidenstein, S. and Goddard, C. J. (1980). Human factors survey of electrical power control centers. In G. Corrick, E. Haseltine, & R. Durst (Eds.), *Proceedings of the 24th annual meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.